

Université de Montréal

A Socio-Technological Approach to Sharing Knowledge Across Disciplines

par

Sébastien Paquet

Département d'informatique et de recherche opérationnelle

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en Informatique

Mars 2002

© Sébastien Paquet, 2002



QA

76

U54

2003

V.010



AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:
A Socio-Technological Approach to Sharing Knowledge
Across Disciplines

présentée par:
Sébastien Paquet

a été évaluée par un jury composé des personnes suivantes:

Jean Vaucher
président-rapporteur

Esma Aimeur
directrice de recherche

Gilles Brassard
codirecteur

Marc Kaltenbach
membre du jury

Tommaso Toffoli
examineur externe

.....
représentant du doyen de la FÉS

Résumé

La récente explosion des savoirs et leur fragmentation en une constellation de disciplines et sous-disciplines, chacune avec son langage spécialisé, pose un obstacle particulier au partage de connaissances. Les barrières de langage interdisciplinaires réduisent l'impact des travaux de recherche en entravant la migration des idées hors de leur domaine d'origine. Elles induisent de fréquentes réinventions de la roue et gênent le travail efficace sur certains des problèmes de recherche contemporains les plus importants.

Cette question est d'autant plus préoccupante que nous sommes à l'ère de l'Internet. Il n'y a jamais eu autant de connaissances à portée de la main; or il demeure extrêmement difficile de localiser les ressources aptes à satisfaire nos besoins en connaissances.

Dans cette thèse, je conceptualise la problématique du partage interdisciplinaire de connaissances comme une conjugaison de problèmes socio-culturels et technologiques, et je décris les contraintes qu'une trajectoire nous sortant de cette situation devrait satisfaire pour être pratique.

Je propose et analyse ensuite une trajectoire possible pour faciliter le partage de connaissances entre les disciplines. Cette trajectoire comprend trois technologies interreliées basées sur le World Wide Web, qui sont autant d'étapes vers une organisation plus satisfaisante des gens et des connaissances. Ces technologies sont: la *publication personnelle de connaissances*, les *répertoires de connaissances ouverts et partagés* et les *ontologies de synthèse navigables*.

J'évalue l'efficacité des deux premières technologies sur la base d'études de cas et d'une enquête menée auprès de gens qui ont utilisé ces outils. J'analyse aussi mon expérience personnelle dans la construction et l'utilisation d'ontologies de synthèse navigables.

Les résultats de ces évaluations indiquent que la publication personnelle de connaissances permet effectivement une circulation plus aisée des connaissances entre les disciplines, en permettant à des liens forts de se développer entre personnes qui œuvrent dans des domaines différents. Ils suggèrent aussi que les répertoires de connaissances ouverts et partagés sont une bonne façon d'organiser des communautés de recherche et des corps de connaissances autour de thèmes interdisciplinaires. Enfin, les ontologies de synthèse navigables apparaissent comme un médium approprié pour lier des éléments de connaissance qui sont superficiellement différents mais portent la même structure profonde, facilitant ainsi le transfert d'idées d'une discipline vers l'autre.

Mots clés: Communication interdisciplinaire, partage de connaissances, représentation des connaissances, architecture des connaissances, concepts unificateurs, réseautage académique, réseaux sociaux, émergence sociale, systèmes ouverts d'hypertexte, communautés virtuelles.

Summary

The recent explosive growth of knowledge and its fragmentation into a constellation of fields and subfields, each with its specialised language, poses particular obstacles to knowledge sharing. Interdisciplinary language barriers reduce the impact of research work by preventing it from migrating across fields. They induce frequent reinventions of the wheel. They also hinder effective work on some of the most pressing contemporary research problems.

The issue is especially tantalizing in the current era of the Internet. Never before have we had so much knowledge available at our fingertips; yet it remains extremely difficult to pinpoint and locate resources apt to satisfy our knowledge needs.

In this thesis I conceptualise the problem of interdisciplinary knowledge sharing as a conflation of socio-cultural and technological problems, and I describe constraints that a successful solution path out of this situation ought to satisfy to be practical.

I then propose and analyse a path towards enabling effective knowledge sharing between disciplines. This path is comprised of three interrelated World Wide Web-based technologies that represent successive steps towards a more coherent organisation of people and knowledge. These technologies are: *personal knowledge publishing*, *open shared knowledge repositories*, and *navigable synthesis ontologies*.

I assess the effectiveness of the first two technologies based on case studies and surveys of people who have used such tools. I also report on my personal experience in building and using an interdisciplinary ontology.

The results of these assessments indicate that personal knowledge publishing does in practice enable knowledge to flow more easily across disciplines by enabling

strong relationships to be built between people who specialise in different areas. They also suggest that open shared knowledge repositories are a good way to organise research communities and bodies of knowledge around interdisciplinary themes. Finally, navigable synthesis ontologies appear to constitute an appropriate medium for linking together pieces of knowledge that are superficially different but carry the same deep structure, thus facilitating the transfer of ideas from one discipline to another.

Key words: Interdisciplinary communication, knowledge sharing, knowledge representation, knowledge architecture, unifying concepts, scholarly networking, social networks, social emergence, open hypertext systems, online communities.

Contents

Résumé	iii
Summary	v
Table of Contents	vii
List of Tables	xiii
Table of Figures	xv
Acknowledgements	xix
1 Introduction	1
1.1 The interdisciplinary knowledge sharing problem	2
1.1.1 The story of Jim	2
1.1.2 The general case	4
1.1.3 Interdisciplinary language barriers	5

TABLE OF CONTENTS

viii

1.1.4	Effects of interdisciplinary language barriers	7
1.1.5	Dissecting the problem	10
1.1.6	Constraints for adequate solutions	12
1.2	Contributions of this work	13
1.3	Structure of the thesis	14
2	Related Work	15
2.1	Scholarly communication	15
2.1.1	Scholarly communication in article form	16
2.1.2	Other forms of scholarly communication	18
2.1.3	Interdisciplinary knowledge sharing and the traditional publication system	22
2.2	Knowledge Management	23
2.2.1	Knowledge communities	24
2.2.2	Ontologies	25
2.3	Cognitive Science	27
2.3.1	Communication and learning	27
2.3.2	Analogy and metaphor	27

3	Personal Knowledge Publishing	30
3.1	The need for personal connections	31
3.2	Defining “personal knowledge publishing”	32
3.2.1	Weblogging	32
3.2.2	Personal knowledge publishing	39
3.3	An illustrative example	40
3.4	Implementations of personal knowledge publishing	44
3.5	Theoretical considerations: how personal knowledge publishing works	45
3.5.1	Exploring	47
3.5.2	Pushing	48
3.5.3	Pulling	49
3.5.4	Regular reading	49
3.6	Limitations of personal knowledge publishing	51
3.7	Suitability for interdisciplinary work	53
4	Open Shared Knowledge Repositories	56
4.1	The need for open collaboration	57
4.2	Defining the term	60
4.3	An illustrative example	63

TABLE OF CONTENTS

x

4.4	Implementation of open shared knowledge repositories	66
4.4.1	Traditional Wiki	66
4.4.2	Individual Custody Wiki	66
4.5	Theoretical considerations: how open shared knowledge repositories work	67
4.5.1	Using the repository to represent and find knowledge . . .	67
4.5.2	The body of knowledge as an attractor	68
4.5.3	The body of knowledge as community glue	70
4.5.4	The body of knowledge as foundation for new knowledge .	72
4.6	Limitations of open shared knowledge repositories	73
4.7	Suitability for interdisciplinary work	76
5	Navigable Synthesis Ontologies	78
5.1	The need for navigable synthesis ontologies	79
5.2	Defining the term	81
5.3	A short example	82
5.4	Implementation of navigable synthesis ontologies	86
5.5	Theoretical considerations: how navigable synthesis ontologies work	88
5.5.1	Using the ontology to locate knowledge	88
5.5.2	Growing the ontology	91

TABLE OF CONTENTS	xi
-------------------	----

5.5.3 Using the ontology to locate knowledgeable people	94
5.6 Limitations of synthesis ontologies	95
5.6.1 Limitations of my implementation	95
5.6.2 Limitations of the general design	96
5.7 Suitability for interdisciplinary communication	97
6 Assessing the effectiveness of the tools	99
6.1 Notes on methodology	100
6.2 Personal knowledge publishing	101
6.2.1 Case studies	101
6.2.1.1 Lawrence Lessig	101
6.2.1.2 Lilia Efimova	105
6.2.1.3 Sébastien Paquet	108
6.2.1.4 Analysis	118
6.2.2 User survey	119
6.2.2.1 Questionnaire	120
6.2.2.2 Results and analysis	121
6.2.3 Evaluation criteria	132
6.3 Open shared knowledge repositories	134

TABLE OF CONTENTS	xii
6.3.1 Case studies	134
6.3.1.1 Information architecture wiki	134
6.3.1.2 Weblog kitchen	136
6.3.1.3 Analysis	138
6.3.2 User survey	139
6.3.2.1 Questionnaire	140
6.3.2.2 Results and analysis	141
6.3.3 Evaluation criteria	151
6.4 Navigable synthesis ontologies	153
6.4.1 Case study	153
6.4.2 Evaluation criteria	161
7 Conclusion	163
A Cast of Characters	174
B Glossary	176
Bibliography	177

List of Tables

6.1	Distribution of answers to question #7, “Did reading weblogs help you find people whose opinion you would be willing to trust on certain matters of professional interest?”	123
6.2	Distribution of answers to question #16, “Has your blogging activity helped you find people with professional interests close to yours but with a different professional background than yours?”	125
6.3	Distribution of answers to question #17, “ If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through weblogs?”	126
6.4	Distribution of answers to question #19, “How often does weblogging help you fruitfully exchange ideas with people who specialize in a different area than your own?”	128
6.5	Comparison of researchers and non-researchers on the number of interdisciplinary connections made through weblogs.	132
6.6	Distribution of answers to question #7, “Have wikis helped you find people with professional interests close to yours but with a different professional background than yours?”.	142

6.7	Distribution of answers to question #8, “If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through wikis?”	144
6.8	Distribution of answers to question #11, “Based on your experience, would you say that a wiki enables the formation of a community of people with an interest in an interdisciplinary topic more easily than other available means of communication?”	145
6.9	Distribution of answers to question #6, “Are there wikis that you regard as being among the best sources of knowledge on a particular interdisciplinary topic?”	147
6.10	Comparison of researchers and non-researchers on the number of interdisciplinary connections made through wikis.	151
7.1	Comparison of the tools.	169

List of Figures

3.1	Front page snapshot from the weblog of Dave Winer, CEO of UserLand Software, inc., taken on September 4, 2002.	35
3.2	Front page snapshot from the weblog of Lance Fortnow, a researcher at NEC Laboratories America, taken on February 21, 2003.	40
3.3	Front page snapshot from the weblog of Lilia Efimova, a researcher at Telematica Institut in the Netherlands, taken on December 22, 2002.	42
3.4	Front page snapshot from the “Quantum Bits” weblog of this author, taken on December 5, 2002.	43
3.5	Part of the RSS feed for Philippe Beaudoin’s <i>frontierless</i> weblog. .	45
3.6	Radio UserLand’s news aggregator.	46
3.7	Technorati inbound link data for Sébastien Paquet’s <i>Open Research</i> weblog, captured on January 17, 2003. The words in bold characters are links to sites referencing that weblog.	50

4.1	The “Quantum teleportation” page from the free encyclopedia project Wikipedia. The “Edit this page” link (circled) enables visitors to modify the page’s contents.	61
4.2	The page titled “HyperMedium” in Meatball Wiki, and a visualisation of its relationship to neighboring pages.	64
4.3	The revision history for the page titled “HyperMedium” in Meatball Wiki, along with the differences between the last two versions. . .	65
4.4	The backlinks for the “HyperMedium” page in Meatball wiki. (Note that it links to itself.)	65
5.1	Tree concept in a navigable synthesis ontology.	83
5.2	Available instances of the Tree concept.	83
5.3	One possible instance of the Tree concept.	84
5.4	External documents about the Tree concept.	84
5.5	Discussion subpage about the Tree concept.	85
5.6	Unordered tree concept in a navigable synthesis ontology.	85
5.7	Newly created page in my implementation of a navigable synthesis ontology.	87
5.8	Some concepts and their interrelations in a navigable synthesis ontology. Arrows link concepts to their generalisations.	89
5.9	Some concepts and their interrelations in a navigable synthesis ontology. Full arrows denote specialisation; dotted ones denote definitional dependence.	90

5.10	Adding a specialisation called “Palindrome Sentence” to the “Palindrome” concept page.	92
6.1	Lessig Blog, January 18, 2003	103
6.2	Mathemagenic, December 15, 2002.	106
6.3	Seb’s Open Research, first post, July 19, 2002.	109
6.4	Seb’s Open Research, July 29, 2002.	110
6.5	Example of a multi-party conversation on knowledge logging (Mathemagenic, August 20, 2002)	114
6.6	Introduction to the article “Personal knowledge publishing and its uses in research”, Seb’s Open Research, October 3, 2002.	115
6.7	Distribution of answers to question #7, “Did reading weblogs help you find people whose opinion you would be willing to trust on certain matters of professional interest?”	123
6.8	Distribution of answers to question #16, “Has your blogging activ- ity helped you find people with professional interests close to yours but with a different professional background than yours?”	124
6.9	Distribution of answers to question #17, “ If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through weblogs?”	125
6.10	Distribution of answers to question #19, “How often does weblog- ging help you fruitfully exchange ideas with people who specialize in a different area than your own?”	127

6.11 Percentage of the time spent reading for professional purposes that is devoted to reading weblogs vs. number of people with shared interests but with a different background that were found through weblogs. (Square area is proportional to response count.)	131
6.12 The page “Weblog Theory” on Weblog Kitchen.	137
6.13 Distribution of answers to question #7, “Have wikis helped you find people with professional interests close to yours but with a different professional background than yours?”.	142
6.14 Distribution of answers to question #8, “If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through wikis?”.	144
6.15 Distribution of answers to question #11, “Based on your experience, would you say that a wiki enables the formation of a community of people with an interest in an interdisciplinary topic more easily than other available means of communication?”.	145
6.16 Distribution of answers to question #6, “Are there wikis that you regard as being among the best sources of knowledge on a particular interdisciplinary topic?”.	146
6.17 Percentage of the time spent reading for professional purposes that is devoted to visiting wikis vs. number of people with shared interests but with a different background that were found through wikis. (Square area is proportional to response count.)	150
6.18 Part of the page for the “Binary Search” concept in UC Wiki. . .	159

Acknowledgements

First of all I would like to especially thank my advisor, Prof. Esma Aïmeur, without whom this thesis would not exist. She demonstrated genuine interest in what I was doing and was the source of many valuable insights. Her continued guidance and commitment made this journey much more straightforward than it would have been if I had been on my own.

I also wish to thank Prof. Gilles Brassard, my co-advisor, for being an inspirational figure, for believing in me and for supporting my work on a research theme that was fairly off the beaten path.

A huge thank you goes out to research bloggers Lilia Efimova (knowledge management), Sebastian Fiedler (instructional design), Stephen Downes (e-learning), Spike Hall (education), Jill Walker (humanistic informatics), and Thomas N. Burg (knowledge networks) for liberally sharing knowledge with me and amongst themselves.

Also, I wish to acknowledge the moral and intellectual support of like minds such as Chris Corrigan, Flemming Funch, Denham Grey, VeerChand Bothra, Phillip Pearson, and Eric Hanson, none of whom I have yet met but all of whom I now consider friends as well as esteemed colleagues.

Closer to home, I wish to thank many people at Université de Montréal with whom I had many illuminating discussions: Prof. Pierre Poulin, François Duranleau, Mathieu Miller, Sydney Roc, Philippe Beaudoin, Luc Leblanc, Jocelyn Houle, Laurent Lefebvre, and Prof. Alain Tapp.

Finally, I thank my wife, Hélène, and my children, Marie-Jeanne and Étienne, for demonstrating incredible patience in waiting for me at the finish line.

Chapter 1

Introduction

*Les murs qui séparent les départements sont plus épais
que les murs extérieurs des institutions.*

Pierre Dansereau

Everybody laughs in the same language.

Anonymous

In this chapter, I articulate the problem of interdisciplinary knowledge sharing and I describe constraints that a solution to this problem has to satisfy in order to be successful. This provides the essential groundwork and motivation for the rest of the work, which proposes and analyses a path towards enabling effective knowledge sharing between disciplines. This path is comprised of three interrelated World Wide Web-based technologies that represent successive steps towards a more coherent organisation of people and knowledge. These technologies are: personal knowledge publishing, open shared knowledge repositories, and navigable synthesis ontologies.

In Section 1.1, I use a concrete example of how the problem affects research work to introduce the problem. I then define the general problem and analyse it. Section 1.2 states the contributions of this work, and Section 1.3 outlines the structure of the remainder of the thesis.

1.1 The interdisciplinary knowledge sharing problem

In this section, I introduce the central problem that this thesis is concerned with by way of an example, before giving a general description of the problem. Then I analyse the problem and its effects, examine the structure of the problem, and specify constraints that must be satisfied by viable solutions to this problem.

1.1.1 The story of Jim

To introduce the issue I am talking about, I will describe the story of a hypothetical researcher named Jim. The story is specific, but I am really referring to a very generic situation. Throughout the thesis I will refer to Jim's situation to make my points clearer.

Jim is a researcher. He specialises in computer graphics, which is one of the many subfields of computer science. More specifically, he is interested in the problem of global illumination [SP94], in which a physical scene is described mathematically and the goal is to determine the amount of light that is reflected by each surface in the scene.

And yet more specifically, Jim works on a class of algorithms for that problem that are called radiosity methods [GTGB84, CW93, SP94]. Actually, he is trying to find ways of exploiting the geometric structure of scenes for speeding up radiosity algorithms. It's an intriguing problem, one that appeals to Jim because it appears to have some depth to it. But it's a fairly specialised problem.

The first thing Jim needs to do is to find the work by other researchers that relates to this particular endeavor. Jim knows the area of computer graphics pretty well, and he's pretty confident that what he's trying to do has not been done before in that area. But what about other areas? It is quite possible that people in other fields have tackled very similar problems, problems that are perhaps not concerned with light bouncing off surfaces but nevertheless share fundamental characteristics with his problem. As problem-solving pioneer Pólya wrote, "the type of problem suggests the type of solution" [P45].

If such were the case, Jim had better find their papers, because if he doesn't, he just might be poised for yet another reinvention of the wheel. Sure, his work might pass off as original among close colleagues, but he'll have to face major embarrassment if someone suddenly shows up and points out that essentially equivalent work has all been done years ago.

So Jim packs up a lunch and sets out on a quest for prior work. Jim doesn't know other areas very well, and he doesn't know many people in other areas. The few ones he knows are pretty specialised themselves, and they say they can't really give him specific pointers, but they suggest a few areas where he might cast his line and hopefully get lucky: computer vision, computational geometry.

Jim obviously doesn't have the time to take courses in those areas to know the landscape. And even if he did, chances are his problem would not be discussed. So he opts for a computerised search.

As he sits down in front of the computer, it instantly dawns upon him that he does not know the name of what he is looking for. He knows how to describe it in computer graphics-speak, but that won't help him at all in locating material in other fields; he realises he's a tourist in a foreign-language territory.

After trying for a while to navigate the colossal quantity of literature and not finding anything that seemed to quite relate to what he was trying to do, Jim gives up on the search. He sets out on doing the work. He comes up with clever ideas. After a few months' work he has successfully implemented a system to put his ideas to the test. He writes up a paper about it. He then submits the paper to a peer-reviewed journal.

The paper is rejected. The reviewers acknowledge that it is technically accurate and that the results are good. But, as it turns out, one of the reviewers knew someone who is quite knowledgeable in the area of numerical analysis. He called him up and was able to find a couple of papers from ten years ago that did almost exactly the same thing.

Jim is tempted to pull his hair off, but he acknowledges that there was not much that he could do to foresee and prevent this. The prior papers in question were pretty obscure and, although the ideas had the same structure, the authors didn't use the same vocabulary. So finding them was not a trivial matter.

And so Jim sets out on other work, hoping that this unfortunate situation will not happen again. (Or that, if it should happen again, he'll be luckier and encounter less knowledgeable reviewers.)

1.1.2 The general case

The generic scenario that I am considering in this work is the following. A researcher wants to solve some problem that has not been worked on in his area of specialisation. The first thing he needs to do is to find related work. As long as he looks only in his area, this is relatively easy. But where other areas are concerned, the researcher is at a loss to locate relevant material.

In many cases it is quite possible that similar problems exist in other areas, that they have been worked on, and that the relevant material is only keystrokes away, thanks to the Internet; but it is equally probable that these problems have been described using a different vocabulary, making it very hard for the researcher to find it.

One of the adverse consequences of this situation is unnecessary duplication of effort. There are other consequences, which I will describe further on.

1.1.3 Interdisciplinary language barriers

What has happened? Why is this undesirable situation so common? In this section I offer a perspective on the reasons underlying the present state of affairs.

To specialise is to concentrate on a particular, necessarily restricted activity or field of study. Specialisation is a way of increasing overall efficiency; one of the better known examples is Henry Ford's introduction of the assembly line for manufacturing cars. By assigning to each member of the production team a single task to be performed on the cars as they passed by slowly down the assembly line, Ford obtained a tremendous gain in efficiency compared to traditional assembly methods [Lan00].

Observing the contemporary world, it is not difficult to discern a massive trend towards the specialisation of humans, which translates into an explosive growth in the number of fields of human activity. Crane and Small reported the existence of 8530 definable knowledge fields in 1987 ([CS92], cited in [Kle98]). It is reasonable to presume that the figure has grown significantly since then.

Specialised languages. Researchers, in order to be able to push the frontier of knowledge, must focus intensely on particular issues and this is how they become specialised. Since different specialties are concerned with different things, specialists in every branch of knowledge develop a particular vocabulary that enables them to communicate efficiently amongst themselves.

The advantage in developing a specialised language is that of conciseness. Expressing a complex, multi-faceted construct in a single word is usually preferable to using everyday language for either of the following reasons. First, an explicit, unequivocal description of the construct in everyday language would be so long as to be unwieldy. Second, it may not even exist because the construct encompasses tacit knowledge that is only shared through common experience [Pol67, Lak87, NT95, BD00, HK02].

As a case in point, consider the fact that mathematical equations used to be expressed in words. Following the development of a symbol system providing shorthand representations for recurring expressions (i.e. x being substituted for the phrase “the unknown”), it became easier to think about complicated mathematical problems. However, at the same time an additional requirement was imposed unto whoever wanted to understand what mathematicians were doing: it became necessary to learn their specialised symbols and their associated meanings.

Language barriers. A language barrier is what we encounter when trying to understand work in another field than the ones with which we are familiar. Language barriers arise because learned people spend the bulk of their time discussing with people who share their specialisation. This enables them to build effectively upon a foundation of shared understanding. The new knowledge they create is described in terms of the knowledge they share, and the result is a pyramid of meaning that naturally becomes increasingly opaque to outsiders.

At first sight it would seem like there is no simple way to dig holes through language barriers; other than to take the time to become learned in the field. I suggest otherwise, for the following reason: humans are able to use analogies between structurally similar concepts as shortcuts towards understanding. To give one example, the similarity between biological viruses and computer viruses enables us to understand one quite easily when we understand the other, by way of conceptual substitution. Analogy works very well to transmit knowledge, even with complex concepts [LJ80]. Great teachers use analogy very fruitfully to convey ideas.

This is one of the key ideas behind much of the work in this thesis. Another key idea is that there are natural affinities between certain pairs of people who do not necessarily have the same background. As it were, it appears as if these people “think in the same way” or use similarly structured languages to talk with themselves. One possible explanation for this is that they have similar tacit knowledge. People with affinity are also apt to experience efficient transmission of knowledge, even when complex concepts are involved and their respective vocabularies are different. One can think of an affinity connection as a tunnel passing under language barriers.

1.1.4 Effects of interdisciplinary language barriers

Here we explore the consequences of the communication breakdown that generally prevails between people in different fields. A more general discussion of communication problems in science has been provided by Drexler [Dre87].

Preventing work on interdisciplinary problems. The first problem caused by language barriers is the obstacle that it poses to effective interdisciplinary work. Many important current scientific and technological challenges relate to several fields at once. They are interdisciplinary in nature. In order to make progress on those problems, researchers with diverse backgrounds must join forces. This requires them to surmount the language barriers that lie between their respective fields.

Breakthroughs often result from interdisciplinary dialogue, as has been the case most recently with the emergence of fields such as bioinformatics, computational linguistics and quantum information processing. “The ‘inter-discipline’ of today is the ‘discipline’ of tomorrow.” ([OEC72], quoted in [Col02a]).

Moreover, as much research advances towards the study of complex systems [Vic02], whose problems are often ambiguously defined, there is a growing need to find useful parallels between problems that relate to different fields in order to make further progress and better collaborate in thinking about these problems.

Work on interdisciplinary problems is further hindered by the fact that there is often no proper place to report on such work. For instance, most recent papers in network science (e.g. [BA99, BA00, BB01, BJR⁺02]) have been published in physics journals, although they arguably don’t primarily have to do with physics. While it is often the only avenue, publishing interdisciplinary work in a journal that is associated with a particular field is not really appropriate because it is

not addressed to the appropriate group of people. Also, it is typically difficult to do because disciplinary reviewers may be uncomfortable with approving material that does not directly relate to the journal's subject matter.

Reinventing the wheel. The second issue has to do with knowledge reuse across fields. As we have already argued, it is often the case that similar or related problems occur in different domains, which means that there is potential for knowledge reuse in the form of cross-fertilisation. However, such reuse does not occur anywhere near the scale that could be hoped for, especially in an era of mass connectivity with so many ideas being released into the public domain by researchers. This is because there is currently no easy way for a specialist in one field to locate the ideas and developments occurring in other fields that could be of interest to him.

Currently, cross-fertilisation mainly occurs either when a particular idea or achievement is given wide exposure throughout the scientific community, when a specialist migrates into another field, or by chance (such as when researchers with related lines of thought happen to meet in casual settings or social events). None of these happen very often. As a consequence, very few ideas ever make it out of the community they originate from; some are even forgotten; and, sadly, wheels are reinvented on a regular basis. This adds mostly redundant material on top of a mountain of publications, exacerbating the information overload problem already present to a frightening degree.

Obviously, whenever research work is unknowingly duplicated, the original is effectively denied the impact it could have had, and credit is consequently misattributed. For instance, the algorithm for finding minimum spanning trees was referred to as Prim's algorithm for many years, until it was found out that Prim had rediscovered a result found twenty-six years previously by Jarnik [Jar30]. Un-

recognised duplication is a significant concern at the personal level, but a tragedy on the scale of science.

The threat of irrelevance. The third problem caused by the fragmentation of knowledge is easily seen by considering the growth in the number of fields of knowledge.

Each discipline commands an ever-diminishing proportion of the total knowledge available. In such a context, it becomes less and less tenable to ignore what is being accomplished in other communities, as it implies a kind of disconnection from the common reality. A field that is closed upon itself sooner or later finds itself producing arcane, impenetrable material, perhaps quite clever in its own way but essentially unusable by the rest of the world.

By this I do not mean to say that arcane research is to be discouraged; indeed many important pieces of knowledge have resulted from obscure inquiries. But it seems important to ensure that the better part of academic knowledge is not kept inside a plethora of thick, high and unconnected silos [Col02b]), if only to convince people outside academia to continue sponsoring its growth.

I also do not mean that specialisation should be abolished. Specialisation is extremely useful to the growth of progress. I am simply arguing that it is useful to properly deal with some of its negative side effects, and that linking between research communities becomes more and more crucial in that respect.

1.1.5 Dissecting the problem

Here I wish to articulate as clearly as possible the problem I'm concerned with, and what it would mean to solve it satisfactorily. Bearing in mind that this is a

Ph.D. thesis, I would also be well-advised to ensure that the reader will not see the problem as trivial and its solution obvious. Note that to some extent these are contradictory goals, because conceptualizing a problem in the most appropriate way has precisely the effect of making its solution appear obvious [P45].

Recalling the story of Jim, our problem may be formulated as follows: to find new ways of helping Jim efficiently find the knowledge in other fields that relates to the particular question of exploiting geometric structure in radiosity algorithms, given that he does not know the name of what he is looking for.

Obviously we are not only concerned with Jim, the computer graphics specialist; we are concerned with all researchers regardless of their area. From a bird's eye view, the problem is to engineer a system that enables any researcher to better find knowledge that relates to his or her specialised work outside his or her area of specialisation.

Our description of the problem is not complete until we specify the resources that are available to solve it. The resources that are available to us include the following: the communication infrastructure of the Internet; the computing power of the machines that are connected to it; the wealth of documents that are available on it; the researcher himself; and a most critical but often overlooked component, the people who are connected to the Internet.

Knowledge actually lies in the last three components: the available documents, the researcher, and people who use the Internet. Moreover, it is important to recognise that the better part of the total knowledge in the system exists in tacit form in the last component [Pol67, HK02]. People intuitively acknowledge this and act in accordance with it every day when they ask other people for advice before trying to find or invent what they need on their own.

The problem could be seen from two perspectives, a technological one and a

social one [BD91]. From the technological standpoint, knowledge as expressed in documents and information resources is not well-organised enough to satisfy queries of the kind we are considering. From the social standpoint, people are not well-organised enough, which makes it hard to establish pathways from the people who need a particular bit of knowledge to the people who have it.

In reality these issues are deeply intertwined. *Finding what* and *finding who* are intimately tied. Documents refer to one another, forming a complex interconnected network; and documents also have authors. People know about many documents, and they also know many people. So people seeking knowledge frequently use a combination of documents and people to find the knowledge they need [Mor02].

Improving on the current state of affairs means to build, using the resources enumerated above as raw materials, a system where one can navigate the sea of documents and people to serve one's knowledge needs adequately. Such an improvement would obviously also have the reciprocal effect of ensuring that people and documents are found by the people who need them and have greater impact.

1.1.6 Constraints for adequate solutions

The reader will readily acknowledge that we are faced with a daunting task, but it should also be obvious that the resources that are potentially available to face the challenge are not insignificant either.

As was noted previously, any realistic solution directly involves people, and the way it does it is by providing knowledge-sharing tools to people.

To be successful, an interdisciplinary knowledge-sharing tool should exhibit the following characteristics, several of which have been identified by Rogers [Rog95]:

Accessibility. It must be accessible and open to most researchers. A solution that is too costly to implement, either in financial terms or in terms of time investment, will not succeed.

Appearance of value. Its value must be (made) apparent to many researchers, otherwise the tool will not be tried. The presence of a *network effect* (each new user increases the value of the tool) means that the challenge lies more in reaching a reasonable number of early adopters, after which it's mostly downhill.

Adequacy. Its use must result in a better fulfillment of the user's knowledge needs than currently available means, otherwise it will not be used.

Inclusiveness. As much as possible, its use must enable order to emerge without sacrificing diversity by excluding particular areas.

Scalability. It must be apt to scale to large numbers of people. We are faced with a many-to-many communication problem. A solution that breaks down past a few hundred participants is not viable.

Evolvability. As much as possible, the tool must be able to adapt itself to changes in the ways knowledge is represented and accessed. Such changes are to be expected, and a tool that cannot adapt will fail.

Each of the tools that we propose will be assessed with respect to these criteria.

1.2 Contributions of this work

This work makes the following four contributions to the advancement of knowledge:

1. It offers a definition of the interdisciplinary knowledge sharing problem.

2. It describes a World Wide Web-based toolset for facilitating knowledge sharing across disciplines that satisfies the constraints that we have just described. Three tools are defined: personal knowledge publishing, open shared knowledge repositories, and navigable synthesis ontologies. While the first two tools are not new, their application to interdisciplinary knowledge sharing had not been considered previously.
3. It provides theoretical analyses of how these tools should help overcome interdisciplinary language barriers.
4. It empirically evaluates the usefulness of the proposed tools to that end. This is done by way of several case studies and surveys of users of these tools.

1.3 Structure of the thesis

This thesis is divided in seven chapters. This introduction has described the problem that this work is concerned with and has identified specific goals. Chapter 2 introduces earlier research that relates to this work in one way or another. Chapter 3 defines and describes personal knowledge publishing and explains how it contributes to solving the interdisciplinary knowledge sharing problem. Chapters 4 and 5 likewise define and describe open shared knowledge repositories and navigable synthesis ontologies and explain how they help communicate knowledge across disciplines. Chapter 6 provides an assessment of the effectiveness of these tools for sharing knowledge across disciplines. Finally, Chapter 7 summarises my contributions and gives directions for future work.

Chapter 2

Related Work

Whenever I draw a circle, I immediately want to step out of it.

R. Buckminster Fuller

In this chapter, I bring together threads of existing research that relate to the problem of interdisciplinary knowledge sharing. I introduce notions to which I will be referring subsequently in the thesis.

Section 2.1 explains the current state of development of scholarly communication. Section 2.2 describes relevant previous work in the area of knowledge management. Finally, Section 2.3 outlines earlier work in the area of cognitive science that relates to the developments presented herein.

2.1 Scholarly communication

With the rise of the Internet, there has been an increase in research activity around the topic of *scholarly communication*. The better part of the literature

devoted to this topic focuses on formal communication, i.e. communication in the traditional form of academic articles. The issue of electronic journal publishing is at the forefront of these discussions.

Comparatively very little material is concerned with informal scholarly communication, although a crucial part of scholarly communication is not carried out over the medium of formal papers.

In what follows, I first outline what is happening in the area of scholarly communication in the article form. I then provide an overview of new forms of scholarly communication that are emerging outside this arena. I finally discuss what these developments mean for interdisciplinary knowledge sharing.

2.1.1 Scholarly communication in article form

For several centuries now, the article form has established itself as the standard way for researchers to disseminate research results and obtain public credit for their contributions to a shared body of knowledge. The article form as we know it today is the result of a long evolution that began with the records that were written up following the meetings of the early learned societies and circulated among its members, and continued with the advent of the first peer-reviewed journals in 1665 [Wel99].

Until recently, articles were solely distributed in print form, but the computer and networking revolution now makes it possible (and indeed, easier) to distribute articles in electronic form over the Internet.

Making scholarly journal articles available on the Internet as “e-prints” has many benefits for the academic community, including reduced costs, instantaneous access, better ways of navigating the literature, wider and quicker dissemination,

and higher impact of research [Law01, HWB⁺02]. The impetus for a shift to electronic dissemination is quickly getting stronger with the swift rise in both subscription prices and the quantity of published material, which has resulted in a phenomenon known as the “serials crisis”, in which most research libraries have to cancel journal subscriptions every year to avoid budget overflow [Mas02].

Several initiatives have aimed to enable articles to be distributed over the Internet. One of the most successful thus far is Paul Ginsparg’s ArXiv.org ¹, which was set up in 1994 and uses the Web to distribute preprints free of charge in physics, computer science, and mathematics².

The Open Archives Initiative (OAI [RF02]) generalises the ArXiv model and proposes a framework for a universal e-print archive that could be accessed from anywhere. In this framework, many archives distributed all over the world are accessed from one location as if it were a single site. More than one hundred repositories (including ArXiv.org) now conform to the OAI standard.

However, considerable obstacles stand in the way of the evolution to electronic dissemination of articles. Firstly, many publishers of existing print journals resist it, which is unsurprising as it threatens their current business model. Publishers who move to electronic formats usually constrain the dissemination, for example by introducing a delay between the availability of articles in print and online or by restricting access to subscribing institutions.

Secondly, among scholars there is a general perception that electronic dissemination is in some way inherently inferior to its print counterpart, so there is not a lot of demand to publish in electronic form. The most prestigious journals are in print, which results in an understandable inertia on the part of many authors, considering that publication is typically intimately tied to such important career

¹URL: <http://arxiv.org>

²The archive initially had email as the only interface.

issues as promotion and tenure [DGKN00]. Habits are thus hard to change. The slow evolution of electronic publishing has been further discussed by Odlyzko [Odl97], and a discussion of the underlying political aspects of the issue can be found in Kling, Fortuna and King [KFK02].

Globally speaking, things move slowly because the people who would most benefit, such as researchers at institutions with lesser financial means, generally do not have the means to influence the system. The holy grail of free, universal access to the research literature is still far away, but encouragingly, progress is observed every year.

2.1.2 Other forms of scholarly communication

Arguably, the growth of knowledge would come to a grinding halt if the article form were the only allowed means of communication between scholars. Telephone conversations, mail and email, face-to-face meetings and conferences, are essential ingredients in the knowledge sharing that is the lifeblood of academia. Indeed, peer review itself could not be implemented were it not for the availability of these forms of communication. The Internet enables new ways of sharing research knowledge to appear everywhere, many of which are not tied to any formal process.

Here I review how scholars communicate in these other ways, and describe the changes that are occurring there. The section is organised according to the different purposes for scholarly communication.

Peer review. Peer review is the process by which a contribution by a researcher or a team of researchers is evaluated by other researchers to decide whether it ought to be published. Acceptance by a review committee often confers more

credibility to the work and peer reviewed publications typically play a key role in grant and promotion attribution.

Our discussion of peer review is restricted to what is connected to our work. Accounts of the history and current state of the peer review system are provided in the book “Peer Review in Health Sciences” [GJ99] and in an article by Rowland [Row02].

The current process of peer review has evolved under the constraints associated with the traditional print publication model, notably, limited space, high costs and lengthy communication delays. It is typically characterised as “a flawed system, but the best we’ve got”. Its adequacy has been called into question by several researchers in the last decade (see e.g. [Odl95, Smi97, Ens01]). For instance, medical researcher Richard Smith [Smi97] wrote:

The problem with peer review is that we have good evidence on its deficiencies and poor evidence on its benefits. We know that it is expensive, slow, prone to bias, open to abuse, possibly anti-innovatory, and unable to detect fraud. We also know that the published papers that emerge from the process are often grossly deficient.

Litchfield [Lit02] writes:

My own view is that the peer-review system brings some significant downsides to the ideal of free intellectual exchange. There are clearly indicators that the review system prizes conformity over creativity, making it difficult for truly innovative work to find publication. Moreover, every editor knows that the process is not nearly as objective as one might like, at least in the humanities and social sciences, where reviewers are not replicating experimental results.

The Internet allows for new forms of review and commentary that could prove valuable to the research enterprise [Odl00, Lit02]. For instance, the availability of practically unlimited space allows the peer review process to be documented publicly. It also allows feedback on contributions to be collected and displayed more easily.

There have been proposals and experiments with new ways of reviewing research contributions. Some journals, such as the British Medical Journal [Cur02], the Medical Journal of Australia [GJ99], and the Journal of Interactive Media in Education [SS01], use the Internet to implement forms of interactive peer review.

One intriguing proposition is to have retroactive peer review, consisting in making work available before it is formally reviewed. It has been asked if academics could function happily in such a caveat lector³ environment [Lit02]. Indeed, a continuum of peer review, where various editorial structures promote work and ideas from their source into more generally accepted pool of knowledge, has been foreseen [Odl95, Odl03].

Overall, though, there has not been a lot of experimentation and peer review remains to this day a black box that is “remarkably untouched by the rigours of science” [GJ99]. The reluctance to examine and tinker with peer review is mainly attributable to the close tie existing between formal publication and career advancement [Odl00].

The tools for communication that I introduce in this work take advantage of the economics of the new medium that is the Internet. It is now feasible to make everything publicly available without screening all material beforehand to limit volume and cost (provided that one let go of the assumption that published material is necessarily quality material). Approaches that allow this lie at the

³Meaning: “Reader beware”.

opposite end from the traditional peer review system, as they do not use pre-publication control but rather rely on review or filtering happening after the fact.

Shared databases or dynamic resources. The introduction of computers into many research areas has resulted in the generation of considerable information and knowledge that researchers may want to share but that are not suited to the article form. Examples are computer code, gene sequencing information, mathematical data (such as Neil Sloane's On-Line Encyclopedia of Integer Sequences [Odl02, Slo99] and Plouffe's Inverter [BBP97]).

These new forms of research publication could hardly be implemented practically before the Internet. Their availability means that reviewers and members of the research community can check results and re-analyse data in ways not previously possible.

Informal scholarly communication. Researchers routinely exchange knowledge informally with one another. Personal face-to-face contacts, correspondence, and conferences play an essential part in the overall growth of knowledge [Nar00]. To give just one example, scholars often rely on personal recommendations for guidance in selecting material to look at ([KSS97, Odl02]).

However, because this interaction is generally not publicly recorded, it tends to be taken for granted and has a tendency to become invisible. There is hardly any literature on this topic, and it is difficult to research it as most informal communication is not public.

The Internet obviously provides new ways to conduct informal communication. One important question is whether these tools will set up new dynamics in scholarly communication.

One way it could have an effect is by transforming what is known as the *invisible college*. The invisible college for a given research community is “a social network of generally around 100 individuals who function as the scholarly in-group” who produce most of the research in that area and are aware of new developments before they make it into print [Cro82, Gre94]. Examining the impact of email and mailing lists on informal scholarly communication, it has been predicted that they would have the effect of enlarging the invisible college to a much greater size, and that “the elitism of the old scholarly networks would be overcome” [Gre94].

This prediction does not seem to have materialised yet. One possible reason is that the available structures for public electronic communication do not scale to that size. It is often observed that the presence of many novice members in a group dilutes the discussion. As a result, scholars on the cutting edge are likely to default to private channels to avoid noise in their discussions.

Informal communication is doubly crucial where interdisciplinary work is concerned. This is especially true in the early stages, because interdisciplinary developments often don’t initially fit in formal structures. This is why the first two of the tools that I propose in this thesis have a good degree of informality about them.

2.1.3 Interdisciplinary knowledge sharing and the traditional publication system

Efforts towards interdisciplinary knowledge sharing should not be expected to succeed easily within the traditional publication system, for several reasons.

First, the traditional publication system does not easily lend itself to interdisciplinary knowledge sharing. Remember that the goal is to let knowledge reach

the people who need it. Interdisciplinary work typically does not fit anywhere in the traditional disciplinary structure of scholarly journals. Publishing interdisciplinary work in a disciplinary journal can be attempted, but even if it is accepted it will reach only a fraction of the right population of readers.

It is possible to set up interdisciplinary journals, but these are seldom taken seriously. It is often difficult to assemble a cohesive editorial board and to define a clear focus for an interdisciplinary journal. Even if that were done, the journal itself would not be very useful as a focal point because of its lack of focus. Electronic distribution is usually the only practical possibility, which agains plays against credibility. As a result, few people would consult it.

The argument in the above paragraph does not apply if there has been enough work on a particular interdisciplinary area to enable a critical mass of researchers to cluster and define the area in a sufficiently coherent way. But in this case we have witnessed the birth of a new field, and this is why the disciplinary approach has become suitable.

The traditional publication system has considerable inertia, as I have pointed out previously. Given that the comparatively simple issue of electronic dissemination has only progressed slowly in the last ten years, it appears reasonable to expect that if interdisciplinary knowledge sharing is to flourish in the near future, it will happen outside that system.

2.2 Knowledge Management

Knowledge Management is a domain that is concerned with “the formalisation of and access to experience, knowledge, and expertise that create new capabilities, enable superior performance, encourage innovation, and enhance customer value

[in organisations]” [Bec97]. Overviews of the field can be found in the references [DGD⁺98] and [Lie99].

Knowledge management initially grew out of such areas as artificial intelligence and information technology. The field could be said to have a split personality in that it features one current that focuses more on information and another that puts emphasis on people. These are sometimes referred to as the *IT Track* and the *People Track* [Sve01]. Of course one suspects that understanding both aspects are essential to success in knowledge management.

Knowledge management is very young and there remain huge unresolved questions, especially as regards human and social factors [TKE01]. While research done in the area of knowledge management relates to the problem of knowledge sharing, the overall orientation of the field is towards enterprise knowledge rather than scholarly networks. Moreover, interdisciplinary knowledge sharing does not seem to have been touched upon in this area.

Two concepts that originate in knowledge management are of particular importance relative to the work presented in this thesis: that of *knowledge communities* and that of *ontologies*. They will be presented in the following subsections.

2.2.1 Knowledge communities

One concept from *People Track* knowledge management is quite relevant to our work: that of a *community of practice*, introduced by Lave, Wenger, Seely Brown, and Duguid a decade ago [LW90, BD91] and subsequently substantially developed [Wen98, HKW98, HKW00, WS00, LS01, LP99]. A community of practice can be defined as “a group of professionals, informally bound to one another through exposure to a common class of problems, common pursuit of solutions, and thereby themselves embodying a store of knowledge” ([MF96], cited in [HKW00]). The

interpersonal connections that exist in communities of practice help distribute the knowledge where it is needed.

A good example of a community of practice is described in Orr's study of photocopier technicians [Orr96] in which he observed that technical knowledge is distributed across the social network of technicians, and that it is retrieved using informal communication.

By enabling new forms of communication, the Internet allows new ways of constituting communities of practice. Online communities of practice have been proposed as a way to facilitate knowledge sharing by allowing conversation and organisation of material [Bow02]. The tools that I introduce in this work can be seen as ways to build online knowledge communities. They are communities of *interest* rather than communities of *practice*, precisely because interdisciplinary connections arise between people who do not share a practice (although they might be doing similar thing on a deeper level).

2.2.2 Ontologies

Ontology is a term originating from philosophy, where it refers to a branch of metaphysics that is concerned with the nature and relations of being [McG02].

In recent years, computation-oriented fields such as artificial intelligence and knowledge engineering have co-opted the term and use it in a different sense than philosophers. An ontology is an explicit specification of a conceptualisation [Gru93b] in the form of a collection of named concepts, instances, relations and axioms. Together, these constitute a statement of a logical theory that represents a body of knowledge.

Although ontologies initially emerged as a way of allowing the reuse of domain

knowledge in knowledge-based systems, in recent years there has been a growing interest in ontologies and many additional uses for them have been identified [SBF98]. Among other things, ontologies can be exploited for knowledge management [BFGP98, SSSS01], they promote interoperability between databases and information systems [WG97], and they can be useful in natural language processing [Mil90]. Moreover they are centrally involved in Semantic Web development efforts [BLHL01].

Most applications that use ontologies are purely software systems. However, by providing a shared understanding, ontologies can facilitate communication not only between computer systems, but also between people [UG96].

Other researchers have worked on using ontologies to improve communication between scientists. The (KA)² initiative [BF98, BFDGP99] uses an ontology to make knowledge about members of the knowledge acquisition research community more accessible. Buckingham Shum, Motta and Domingue [SMD00] propose using an ontology-based digital library server to support scholarly interpretation and discourse. Kalfoglou *et al* [KDM⁺01] argue for the role of ontology-driven personalised Web-based services in information sharing.

Work on generic ontologies and problem-solving methods has been successful in characterizing several domain-independent constructs [SAA⁺00, vHSW97]. Such constructs play a pivotal role in my approach because they can serve as bridges between domains, as outlined in Chapter 5.

The need to allow ontology development to proceed in representations better suited to humans than computer code has been recognised [GF92, UG96]. In this perspective, ontology development tools that rely on a set of intermediate representations that more closely resemble the way experts are used to thinking about their domain have been proposed [BFGPGP98, FPJ97]. Finally, a number of tools to enable collaborative ontology development have been developed, e.g.

the Ontolingua server [FFR97], Ontosaurus [SRKR97], Tadzebao and WebOnto [Dom98]. However, these tools are designed to be used by knowledge engineers who wish to build formalised ontologies, and thus are not well-suited to our particular purpose of allowing knowledge sharing between scholars in fields that have little to do with knowledge engineering.

2.3 Cognitive Science

2.3.1 Communication and learning

In order for knowledge to be efficiently assimilated, it is necessary that it take a form that is appropriate to the particular context of the person who needs to acquire it. Thomas [Tho78] argued that a “design-interpretation” model is appropriate to describe this state of affairs. In Thomas’ model, the originator of a message uses knowledge about the context and the recipient to design a communication that will have the desired effect when received. This model is consistent with the observations of theorists such as Dewey [Dew33], Vygotsky [Vyg62], and Piaget and Inhelder [PI69].

In this work, this dependence on the person seeking to build knowledge translates into self-directed exploration and designing knowledge environments that allow multiple entry points into the same pieces of knowledge to accommodate the diversity of knowledge seekers’ backgrounds.

2.3.2 Analogy and metaphor

Analogy is a theme that goes deeply into the nature of human thought. It is the basis for how we regroup things and ideas. It also involved in human reasoning.

Analogy allows us to reuse knowledge from one domain of experience in another domain. Linguist George Lakoff and philosopher Mark Johnson who have perhaps made the most important contribution to our understanding of analogy, wrote that “we understand experience metaphorically when we use a *gestalt* (big-picture view) from one domain of experience to structure experience in another domain.” [LJ80]

Lakoff’s central thesis is that metaphors facilitate thought by providing a framework grounded in experience that allows new concepts to form. Lakoff insists on positioning the cognitive agent as an essential player in this process: his thoughts are structured in a subjective manner that depends on his previous experiences and the ways of framing situations that come from them.

An illustration of this is the fact that spatial metaphors do not work the same way depending on the direction chosen. Objectively, there is no difference between the “forward-back” and the “left-right” directions. It is necessary to take a subjective perspective to see the difference, and this explains why there are more numerous and stronger language constructions that are tied to “forward-back” (e.g. “going back in time”, “stepping forward”) than to “left-right”.

Metaphors are used in psychotherapeutic contexts to facilitate the emergence of new patterns of thought, feelings, and behaviour [Esp01]. It is natural to expect them to enable individuals to attain new understandings in non-therapeutic contexts. It is often overlooked how fundamental a role analogy plays in science. Gentner and Jeziorski [GJ93] write that

“Analogy and metaphor are central to scientific thought. They figure in discovery, as in Rutherford’s analogy of the solar system for the atom or Faraday’s use of lines of magnetized iron filings to reason about electric fields. They are also used in teaching: novices are told

to think of electricity as analogous to water flowing through pipes or of a chemical process as analogous to a ball rolling down a hill. Yet for all its usefulness, analogical thinking is never formally taught to us. We seem to think of it as a natural human skill, and of its use in science as a straightforward extension of its use in commonsense reasoning. For example, William James believed that ‘men, taken historically, reason by analogy long before they have learned to reason by abstract characters’. All this points to an appealing intuition: that a faculty for analogical reasoning is an innate part of human cognition”

The reuse of knowledge hinges upon analogy in conceptual structure. In Chapter 5 of this work where I describe the third interdisciplinary knowledge sharing tool, I propose to use analogy in a systematic manner to organise knowledge.

Chapter 3

Personal Knowledge Publishing

I think that what continues to be exciting about online social networks is that people have the power to build them themselves.

Howard Rheingold

In this chapter, I describe the first of the three tools that I propose to use for knowledge sharing: *personal knowledge publishing*, and I explain how it can be used as a way of fostering interdisciplinary communication. In a nutshell, this activity enables researchers to individually articulate their respective research interests in a public manner and build relationships with other researchers based on shared interdisciplinary interests.

The concept of personal knowledge publishing synthesises observations of how people share knowledge in informal face-to-face settings, and in Web-enabled systems such as weblogs (described in Section 3.2). Although there are many weblogs in existence, there is little research literature on the subject and many basic concepts have not been understood and described explicitly. Hence, one of the goals of this work

is to provide a foundation for discussing these systems and their modes of operation.

Section 3.1 highlights the central role of interpersonal connections in allowing knowledge to cross disciplinary boundaries, providing a specific motivation for the ideas in this chapter. Section 3.2 derives a definition of personal knowledge publishing from a definition of weblogging, and Section 3.3 provides an illustrative example of personal knowledge publishing. Section 3.4 discusses implementations of personal knowledge publishing. Section 3.5 explains how personal knowledge publishing can be used and how it works to allow knowledge to circulate across disciplines. Section 3.6 identifies limitations of personal knowledge publishing with respect to knowledge sharing. Finally, Section 3.7 explains why personal knowledge publishing is especially enabling for interdisciplinary knowledge sharing.

3.1 The need for personal connections

Let us go back to Jim, the researcher from the introduction, to motivate the ideas of this chapter. Recall that, as the first step in digging into his newfound interest into scene structures, Jim needs to find the work by other researchers that relates to this particular endeavor. Also recall that most of Jim's knowledge lies in the field of computer graphics and that almost all of Jim's acquaintances work in that same field. This is a typical situation, as maintaining good contacts with researchers outside one's field is difficult in the absence of social events, shared endeavors or geographic proximity.

The result of this situation is that, in all likelihood, a lot of the knowledge that could benefit Jim is for all practical purposes inaccessible to him. He lacks the

vocabulary to formulate what he could look into; and he lacks the connections to people who could guide him.

Now imagine that Jim had at his disposal the means to build and maintain a personal network of people from many disciplines, with whom he shares research interests, who are on the same wavelength as him (i.e. mutual understanding is easy to reach), and whom he trusts. Then, when the time comes to look for knowledge outside his field, Jim could tap into that network to locate relevant literature and people much more easily.

What I propose in this chapter is for individual researchers to use personal knowledge publishing as a tool for building strong interdisciplinary research networks. Although the tool in itself is not new, its application to interdisciplinary knowledge sharing has not been studied before.

3.2 Defining “personal knowledge publishing”

3.2.1 Weblogging

Definition. Personal knowledge publishing is a particular form of the practice known as *weblogging*. In order to define personal knowledge publishing, I therefore need to first define what a *weblog* is.

Weblogs may be viewed as an evolved form of personal Web pages, or “home pages”. The term, coined by Jorn Barger in 1997 (although the form had been in existence for a while), refers to a web site that is a “log of the Web”, indicating a record that points to material available on the World Wide Web.

Many definitions of the term “weblog” have been proposed (e.g. [Win03, MW02]). Since the genre is evolving quickly, there is currently only a rough consensus on

what properly constitutes a weblog. However, a number of features are commonly agreed to be defining characteristics of the genre. I will use the term weblog to refer to a web site that exhibits those features:

1) *Personal editorship*. The content of the site is under the responsibility of a single person (although visitors may post comments in designated sections) and to some extent reflects this individual's personality. Whereas the creation of web pages may be outsourced, you cannot have someone else run your weblog, because then it would no longer be your weblog.

2) *Hyperlinked post structure*. The site's contents consists in typically short posts that feature hypertext links referencing material outside the site. These may be links to news items from sources such as CNN.com or the New York Times Online, or to other weblog posts. The selection of links is entirely up to the editor, who may link anywhere on the web. There is also no prescribed length for a post – some posts simply consist of a single link to content elsewhere, but most often they also include additional information and/or personal commentary on the issue under discussion. The presence of links is what distinguishes the weblog from the online diary, in which an author mostly recounts personal events and thoughts, and which is not especially relevant to anyone outside the author's circle of friends. Each post has a *permalink*, that is, a URL that leads specifically to it.

3) *Frequent updates, displayed in reverse chronological order*. A weblog is a continuously-running publication, much like a daily or weekly newspaper. The latest posts (hence the freshest content) appear at the top of the weblog's main page, and older content appears further down. New posts typically appear from several times a month to several times a day. This characteristic creates an expectation of updates that incites readers to visit the site on a regular basis. A relationship is established between author and reader and strengthened with each visit, just as happens with other regular publications. This probably marks the

most fundamental distinction between weblogs and personal “home pages”, the latter often being seen once and seldom revisited.

4) *Free, public access to the content.* The site’s content is freely accessible via the World Wide Web without restriction such as payment or membership. (This is often taken for granted on the Web, but it distinguishes weblogs from commercial forms that make sharing more difficult.)

5) *Archival.* While older posts may disappear from the front page, they are archived and may be accessed elsewhere on the site. Each post is assigned a permanent hyperlink or permalink which makes it possible to reference older material.

The related term *K-log* or *Knowledge log* is often encountered on the Web. It refers to a generalisation of personal knowledge publishing. The criterion that is removed is the fourth: some K-logs may not be available on the Web, but rather to a closed group such as the employees of a particular private company. For instance, the employees of UserLand use K-logs to circulate knowledge around the company. A completely private (e.g., password-protected) K-log may also be useful to take notes on one’s work.

Brief history. The first weblog was Tim Berners-Lee’s “What’s New?” page¹, which pointed to new Web sites as they came online, starting in 1992. The second weblog was Marc Andreessen’s “What’s New?” page at the National Center for Supercomputing Applications, which performed a similar function until mid-1996.

Several new weblogs appeared with the explosion of the web in 1996-1997. Early weblogs include Dave Winer’s Scripting News² (see Figure 3.1), Jorn Barger’s

¹URL: <http://info.cern.ch/>

²URL: <http://scripting.com>

Robot Wisdom³, and Cameron Barrett's CamWorld⁴. Although it is now collectively edited, Rob Malda's Slashdot⁵ deserves mention, as it became (and to this day remains) phenomenally popular.

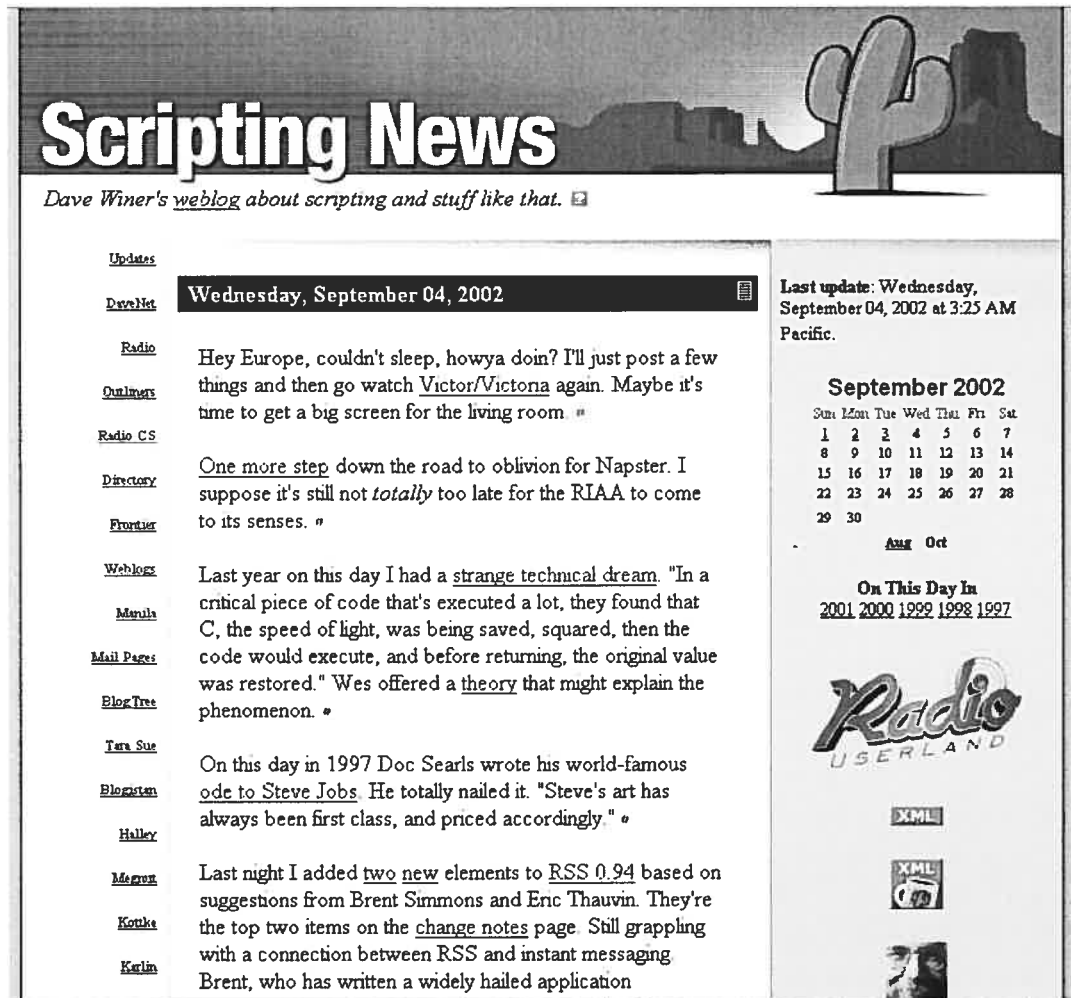


Figure 3.1: Front page snapshot from the weblog of Dave Winer, CEO of UserLand Software, inc., taken on September 4, 2002.

The content of early weblogs was most often a mix of links and commentary that was tailor-made to their editor's taste. Over time, some of those weblogs built sizable followings because they provided a unique selection of fresh content that

³URL: <http://robotwisdom.com>

⁴URL: <http://camworld.com>

⁵URL: <http://slashdot.org>

appealed to a segment of the online population and because of their personal flavour.

Most of the first weblogs were home-grown by web designers and software developers, who were the most aware of the capabilities of the technology. In the early years, there existed only a handful of them. In 1999, however, several free or inexpensive weblogging services, such as Pitas⁶, Livejournal⁷, Pyra Labs' Blogger⁸ and UserLand's EditThisPage.com⁹, were introduced. As using these systems requires little technical knowledge, the practice suddenly became much more accessible. This resulted in a spectacular growth in the number of weblogs.

There were around a thousand weblogs in mid-2000. As of mid-2002, estimates put the number of weblogs at around a half-million, with the Blogger system alone currently reporting more than 350,000 registered users and creating a new weblog every 40 seconds, or more than 60,000 a month [Lev02]. Not all weblog editors write daily, but weblogs are generally considered inactive when they are updated less than once a month. Typical active weblogs are updated a few times a week.

One of the most significant things that happened with the growth of the weblog community is that weblogs became a conversational medium. Many editors would use their weblog to discuss things that had been said by another editor, using links to enable readers to follow threads. Arbitrary numbers of people could participate in such conversations, provided they had their own weblog.

On the surface, this may appear like another incarnation of online many-to-many communication as already implemented in the form of newsgroups or mailing lists. However there is a crucial difference. Since contributions are posted on their author's space, replying to someone else's post does not necessarily mean

⁶URL: <http://pitas.com>

⁷URL: <http://livejournal.com>

⁸URL: <http://blogger.com>

⁹URL: <http://editthispage.com>

that the reply will be seen. This has a bearing on the quality of the material that one can find in weblogs, as I will explain shortly.

Another thing that happened during the recent boom was that, as webloggers started reading other people's weblogs, a practice called *blogrolling* became widespread. The Microcontent News glossary [Hil03] defines a blogrolling list as "The section of a weblog that lists the sites that the blogger reads on a regular basis. This is usually located on the side of a blogger's frontpage, or on a separate page linked off of the frontpage." A unique aspect of these link lists is that they make explicit the social connections that exist among webloggers.

In recent years, the weblog phenomenon has continued its expansion, and the population of webloggers has become increasingly diverse. Growing numbers of professionals have started weblogging and use them "to reflect upon their work, to follow developments in the field, and to publish ideas" [MW02]. Apart from software developers and web designers, the most well-represented professions in the "blogosphere" are information architects, journalists (starting in the fall of 2002, UC Berkeley is offering a journalism course on weblogging), librarians, lawyers, and education specialists. Knowledge management specialists, information technology consultants and researchers are also increasingly using the medium to engage in conversations about the problems they are trying to solve in their work.

How weblogs foster quality. At this point the reader might wonder, "If there are no reviewers and anyone can write anything in their weblog, how is it then possible to find high-quality content in weblogs?". The answer is that quality emerges in weblogs largely as a result of the web of hyperlinks that is weaved by the community of editors. Although it is true that there is no review process prior to publishing, one definitely occurs immediately after publication.

As people read others' weblogs, they link selectively to the content that they

find interesting. Content that has been referenced more often directly obtains more visibility. But this effect is amplified by search engines such as Google that rank web pages according to the number of pages that link to them. As a consequence, when people search for a term, the pages that turn up first are the ones considered most relevant or authoritative by the overall community of editors. The relationship between Google and visibility has been further discussed in articles by Doctorow[Doc02] and Hiler[Hil02].

Note that these dynamics mirror those of academic publishing: articles that are cited more often are more visible and are read more. This is useful in two respects: it encourages quality, and it makes it more likely that people will find the most relevant documents. Obviously, in both cases, relying on the number of citations as a reliable indicator of quality requires that one trust the community. Further note that insofar as references yield visibility, they carry to some extent the power to structure knowledge, a connection has been explored from a critical humanist perspective by Walker [Wal02b].

The other factor that helps quality emerge in weblogs is personal ownership. Although webbloggers participate in a community, the contents of a weblog is not a communal space; it is under the sole responsibility of its editor. Now, most people write in order to be read, to engage other minds. As with any publication, the best way to build and keep a readership is to have an output of consistently high quality. As each individual makes personal decisions as to what to read, higher quality weblogs are read regularly by more people, and thus are linked to more often and appear in more blogrolling lists.

Personal ownership makes weblogs different from other electronic forums like mailing lists. In a mailing list, anyone can write anything and it will automatically end up in everyone's mailbox (unless a moderator is introduced; but this kind of centralised management does not scale easily). By contrast, although a weblog

editor may say anything he pleases, in order to be read widely, he or she must build a reputation over time. In the words of David Walker, Weblogs' users don't vote within the site; they vote by choosing the site as a reliable source of guidance. In effect, they say to the site's author: "you make the choices I'd make if I had time" [Wal02a]. Also, as a weblog keeps all of a person's writing over months or years in a single place, people tend to care more about what they put in it.

To summarise the above two points, quality content is found in weblogs in spite of the absence of centralised control because of the continuous post-publication review process and because good contributions can be published, read, and linked without being bundled alongside lesser quality content.

3.2.2 Personal knowledge publishing

In this section I propose a name and definition for an activity that I call personal knowledge publishing, building on the above definition of weblog. Personal knowledge publishing quite simply consists in *an activity where a knowledge worker or researcher makes his observations, ideas, insights, interrogations, and reactions to others' writing publicly available in the form of a weblog.*

A personal knowledge publication is a specialised form of a weblog; the characteristic which distinguishes it from other weblogs relates to the content. While the definition of a weblog places no restriction on the content that appears in the site, that of personal knowledge publishing asks that the content focus on documenting the conscious activity of creating new knowledge.

As a result of this requirement, the content of a personal knowledge publication closely reflects the research interests that its editor is pursuing. This has direct bearing on the ability of the publication to enable meaningful, constructive networking, as I will discuss shortly.

3.3 An illustrative example

Here I illustrate personal knowledge publishing using a real-life example. I will present a more detailed examination of that example in Section 6.2.

Figure 3.2 provides an illustration of personal knowledge publishing. It represents the weblog of Lance Fortnow, who researches computational complexity at NEC Laboratories.

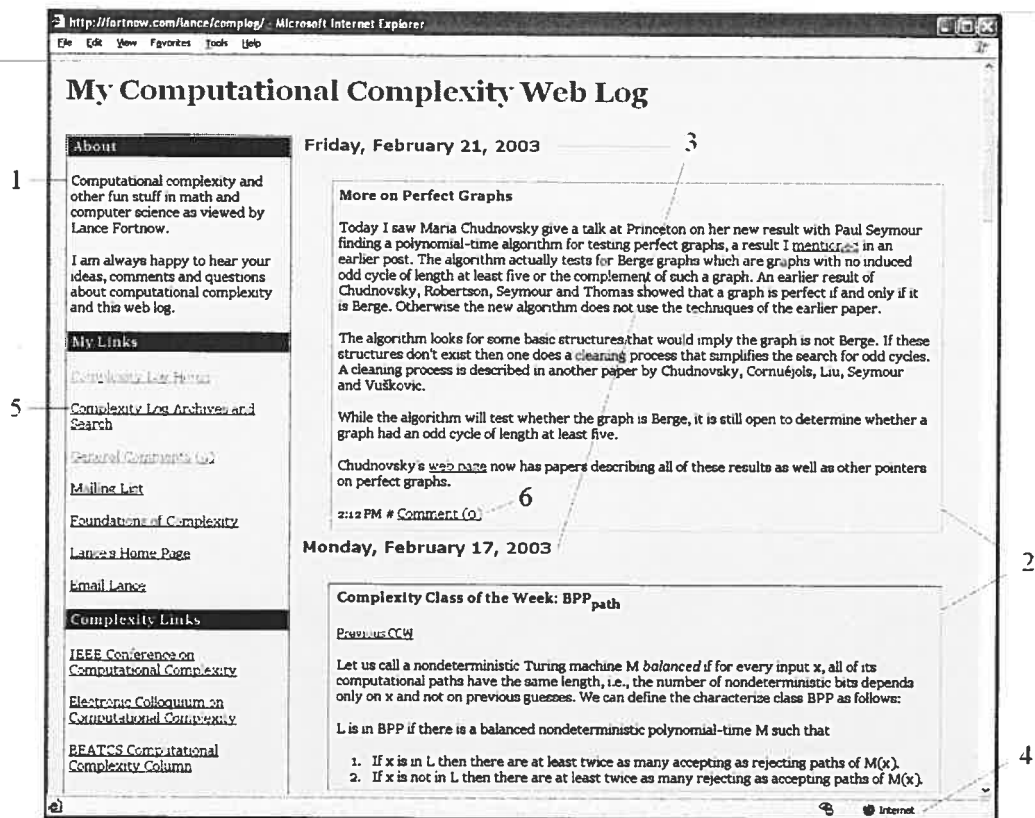


Figure 3.2: Front page snapshot from the weblog of Lance Fortnow, a researcher at NEC Laboratories America, taken on February 21, 2003.

The site presents all the characteristics enumerated in the previous section. The “About” box (1) shows that the site is written by a single person. Two posts (2) are visible on the snapshot (one completely, the other partially), and they feature hyperlinks going both within an out of the weblog. The # sign seen to the right of

the time of the first post is a permalink that enables anyone to reference that post. As seen from the dates, the weblog has been updated twice in the same week (3). Being accessible via the Web at the URL <http://fortnow.com/lance/comprog/>, the site satisfies the free accessibility requirement (4). Chronological structure is evidenced by the dates displayed in the page (3). One of the links in the left sidebar (5) enables access to archive pages for each week that blog posts were written. Fortnow's weblog features subpages where readers may leave comments for the author and other readers (6). The author may reply to them in the same spaces, or in new weblog posts (Private feedback is obviously possible using email).

The visible content is representative of personal knowledge publishing. In the top post, the author discusses an algorithmics problem of interest to him, and links to a web page elsewhere that presents relevant papers and additional links related to the topic. The second post is part of a series that presents various complexity classes for the benefit of readers. Note that the content of the first post can be valuable to the author as well as other people, but would likely be out of place in an article.

It should be noted that Fortnow's weblog is an example of the "regular" use of personal knowledge publishing, that is, discussing a specialized topic of interest to its author. Fortnow is not applying the tool to the problem of making interdisciplinary connections. That this tool can indeed serve to share knowledge *across* disciplines is not obvious from looking at most instances of personal knowledge publishing, and it is one of the questions that my work addresses.

Two more examples of weblogs are given in Figures 3.3 and 3.4.

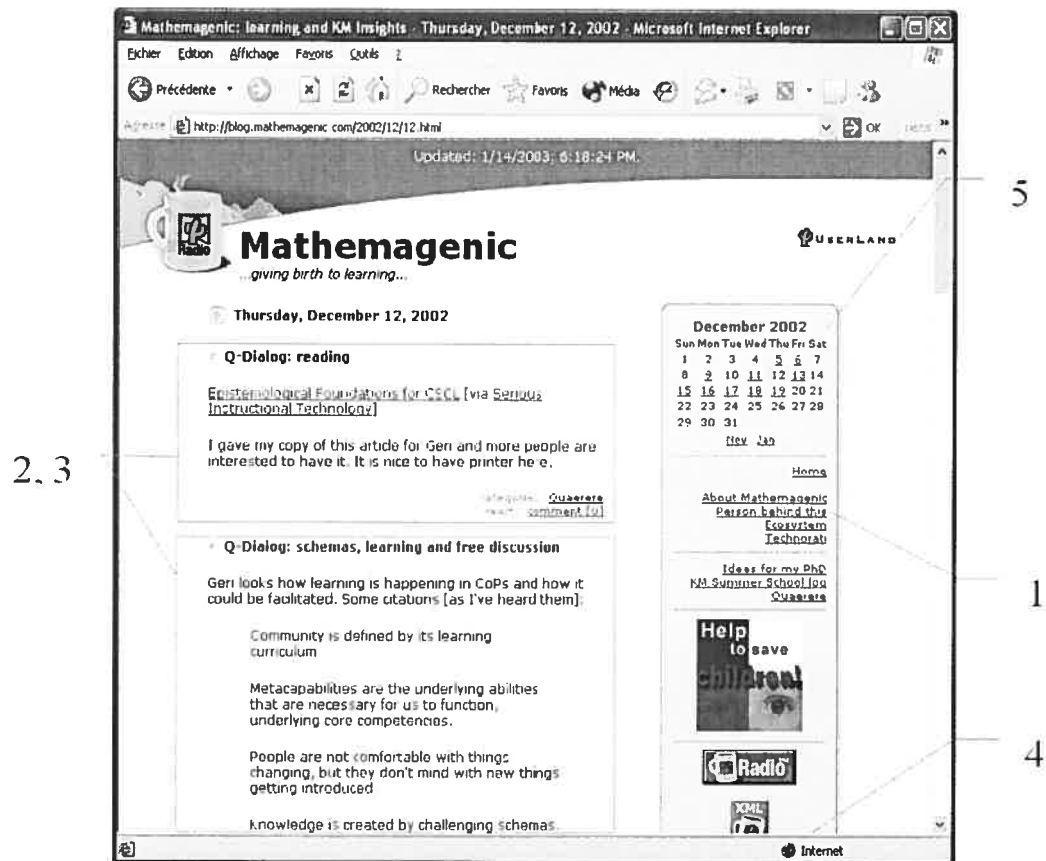


Figure 3.3: Front page snapshot from the weblog of Lilia Efimova, a researcher at Telematica Institut in the Netherlands, taken on December 22, 2002.

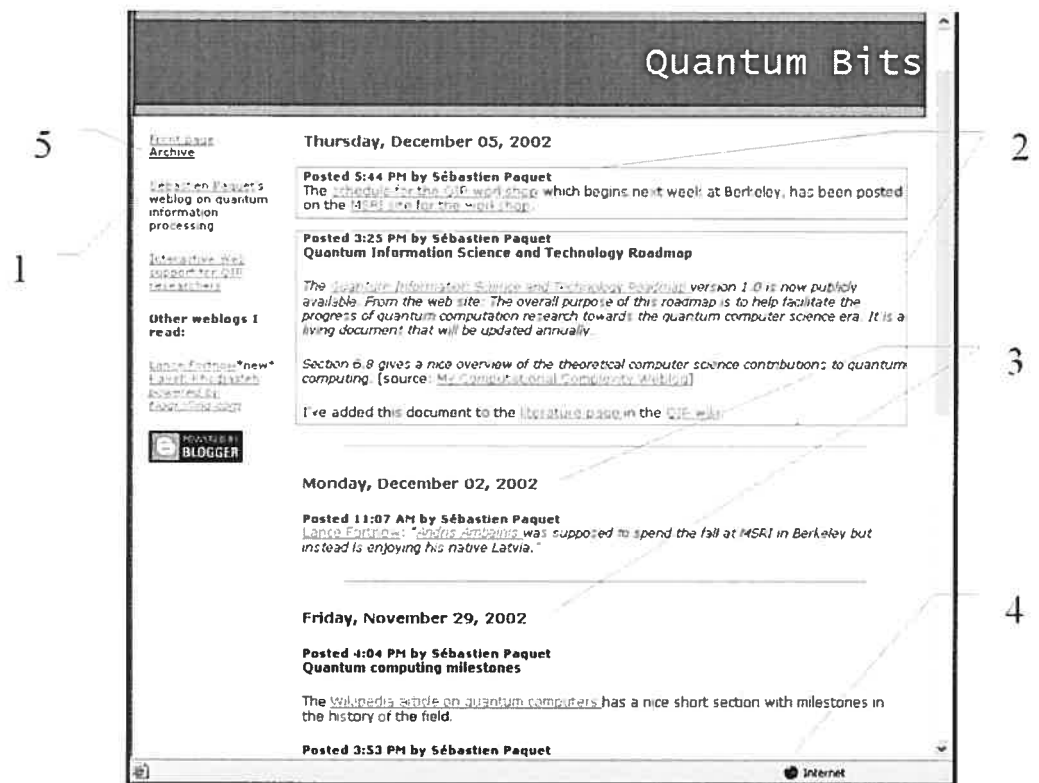


Figure 3.4: Front page snapshot from the “Quantum Bits” weblog of this author, taken on December 5, 2002.

3.4 Implementations of personal knowledge publishing

Several of the weblogging tools that have appeared over the last few years make personal knowledge publishing almost as technically easy as word processing. Among the tools that have all the features required in the definition of “weblog” that was given above, there are notably Blogger, Radio UserLand, and Movable Type.

Weblog tools are typically implemented using a (server-side or client-side) database to store and organise posts and a Web-based interface to handle editing tasks. Publishing is usually done by way of File Transfer Protocol (FTP) to a Web server. Hosting and bandwidth may be provided by the weblog tool company or by the editor himself.

Another feature found on many weblogs is the ability for visitors to leave comments to blog posts, using an HTML form.

In addition to producing HTML, the most advanced weblog tools implement syndication. The idea of syndication in the context of weblogs is to make the content of individual posts available on the web in the Rich Site Summary (RSS) standard format (which is based on the extensible markup language (XML))¹⁰. Figure 3.5 provides an example of a RSS feed.

The combined availability of RSS feeds and software known as *personal news aggregators* makes it possible for readers to select the sources they are interested in and subscribe to them. Subsequently, their aggregator automatically retrieves content from all selected sources and displays them together on their screen, which

¹⁰At the time of writing two different standards are in use.
URLs: <http://www.purl.org/rss/1.0/> and <http://backend.userland.com/rss>

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- RSS generated by RSSify at VoidStar.com, modifications by Philippe Beaudoin -->
<DOCTYPE rss (View Source for full doctype...)>
- <rss version="2.0">
  - <channel>
    <title>frontierless</title>
    <link>http://www2.iro.umontreal.ca/~beaudoin/blog/</link>
    <description>Alternatives to cultural protectionism</description>
    <language>en-ca</language>
    <copyright>Copyright 2002 Philippe Beaudoin</copyright>
    <lastBuildDate>Fri, 17 Jan 2003 11:50:57 -0500</lastBuildDate>
    <docs>http://backend.userland.com/rss</docs>
    <generator>RSSify at VoidStar.com, modifications by Philippe Beaudoin</generator>
    [REDACTED]
    <ttl>60</ttl>
  - <item>
    <title>Jobless</title>
    <link>http://www2.iro.umontreal.ca/~beaudoin/blog/2003_01_12_archive.php#87572704</link>
    <description><br>As of monday I will become temporarily unemployed, waiting for the skies to clear a
    little. If all goes well, I should get my game developer job back within a couple of months. Meanwhile I
    will have plenty of time to spend with my new son and with my weblog. There is a least one experiment
    I've been willing to put up in the <a
    href="http://www2.iro.umontreal.ca/~beaudoin/blog/weblab/">weblab</a>, I should have some time
    to program it now.<br></description>
    <guid
    isPermaLink="true">http://www2.iro.umontreal.ca/~beaudoin/blog/2003_01_12_archive.php#87572704</guid>
    <pubDate>Thu, 16 Jan 2003 23:28:06 -0500</pubDate>
    <comments>http://www2.iro.umontreal.ca/~beaudoin/blog/comments.php?
    2003_01_12_archive.php*87572704</comments>
  - </item>
  - <item>
    <title>Why People Wont Blog</title>
    <link>http://www2.iro.umontreal.ca/~beaudoin/blog/2003_01_12_archive.php#87571977</link>

```

Figure 3.5: Part of the RSS feed for Philippe Beaudoin's *frontierless* weblog.


means that they don't have to actually visit any of these sites to look for new content. Figure 3.6 illustrates the output of a news aggregator. It is possible to subscribe or unsubscribe from any particular source at any time. Many people publish such a feed and use aggregators to make more efficient use of their time.

I won't further dwell on the technical aspects of implementing these tools, because the issue in this work is how a researcher can exploit them to build strong connections with people outside their discipline. This is explained in the next section.

3.5 Theoretical considerations: how personal knowledge publishing works

In this section I describe the process whereby a researcher can build an interdisciplinary network around his personal interests using personal knowledge publishing

News Aggregator

This page lists the most recent stories from the channels you've [subscribed](#) to. Click on the POST button to add a story to your weblog, to categorize the story, or edit its text. You can delete stories from this page. Check the stories that you want to delete, then click the Delete button. 

Delete

















<input type="checkbox"/>	The Motley Fool, 1/10/2002; 10:59:20 AM. 	
<input type="checkbox"/>	Do I Need Long-Term Care Insurance? . There's no one answer that applies to everyone. Here are some helpful criteria.	POST
<input type="checkbox"/>	The Register, 1/10/2002; 10:59:18 AM. 	
<input type="checkbox"/>	IBM relies on SUN workstation for quantum computing . Maybe their own stuff's not good enough	POST
<input type="checkbox"/>	internetnews.com: Top News, 1/10/2002; 10:59:11 AM. 	
<input type="checkbox"/>	A Chemistry Experiment . B2B marketplace ChemConnect is acquiring CheMatch, an Internet exchange for bulk commodity chemicals.	POST
<input type="checkbox"/>	Frontier News, 1/10/2002; 10:59:10 AM. 	
<input type="checkbox"/>	André Radke released a PostgreSQL DLL and a crypto DLL for Frontier.	POST
<input type="checkbox"/>	MacNN, 1/10/2002; 10:59:06 AM. 	
<input type="checkbox"/>	ATI partners with Hercules for Europe	POST
<input type="checkbox"/>	IBM awarded most patents in 2001	POST
<input type="checkbox"/>	BBC News: world, 1/10/2002; 10:59:05 AM. 	
<input type="checkbox"/>	US flies suspects to Cuba base . A first group of al-Qaeda and Taleban prisoners are flown from Afghanistan to a US military base at Guantanamo Bay on Cuba.	POST
<input type="checkbox"/>	The Motley Fool, 1/10/2002; 9:59:06 AM. 	
<input type="checkbox"/>	You Can't Take It With You . What are we all working so hard for? Don't forget money is a means to an end not the end itself.	POST
<input type="checkbox"/>	Disneysoft? . Come on, Microsoft isn't buying Disney. Still, it's fun to imagine the implications.	POST
<input type="checkbox"/>	internetnews.com: Top News, 1/10/2002; 9:59:03 AM. 	
<input type="checkbox"/>	U.S. and U.K. Wireless Ad Groups Merge . The WAA and WMA will become the MMA, and must first cope with different policies on spam.	POST

Figure 3.6: Radio UserLand's news aggregator.

as his principal means of communication. For the process to be successful, it is necessary to assume that other researchers are simultaneously using the same means. This assumption is not a reasonable one to make in most fields as I write these lines, but the current very rapid growth of weblogs among knowledge workers suggests that it will soon be (on a timescale of a few years).

The process is comprised of four activities that I call *exploring*, *pushing*, *pulling*, and *regular reading*. The activities are conducted in parallel.

3.5.1 Exploring

The exploration activity consists for the researcher in surveying the landscape of available content in weblogs to serendipitously find nuggets of particular interest to him. The research-oriented portion of the blogosphere is already too large to consider visiting all research blogs, so a systematic survey is out of the question.

The best way to begin an exploration is probably to read a few weeks' worth of content in a weblog by someone whom the researcher already knows and trusts. This enables one to gather more links to interesting weblogs, along with a bit of context to help decide whether or not to visit them. Links may be found in weblog posts, but also in blogrolling lists. Visiting weblogs one hop away from a trusted weblog provides a wider perspective and allows one to cross-validate perceptions about newly encountered authors.

An alternative way to begin an exploration phase is to use a search engine such as Google for terms describing topics of interest and appending the words "weblog" or "blog" to the query. As those words mostly appear in weblogs, most of the results of the query will point to weblogs discussing the topic in question. From there the strategy described above can then be used.

The result of a successful exploration phase is the identification of a few more new weblogs (and thus authors) of interest, and possibly an improved understanding of one's own interests.

3.5.2 Pushing

The push activity is where the researcher comes out of the shadows. It consists in using one's weblog to actively push information and knowledge out onto the web.

A researcher may push in a number of different ways, depending on the situation. He may link to and comment on a particular web page. He may simply link to a weblog post from another author that he finds relevant; he may complement the link with a quote from that post; and he may react by throwing in his own commentary. To improve the chances of mutual recognition, he may leave a comment on the linked weblog if that is possible, or send email to the author.

Another way to push is to write original posts or longer pieces (usually called "stories"). And yet another way to push is to add a link to someone else's weblog in one's blogrolling list.

In all cases, pushing has the effect of giving a public indication of the researcher's own interest in particular topics or people, enabling others to find him or her in the course of their own explorations. As a weblog inevitably reveals some of its author's personality and quality of thought, visitors can also assess if "wavelengths match" before making a decision to revisit the weblog.

3.5.3 Pulling

The pull activity is where the researcher exploits the responses that his weblog has generated to find yet more people. It boils down to a sort of informed exploration, based on what others found interesting in one's weblog.

The basis for pulling is external links into one's own weblog. Other weblog authors who find content interesting in one's weblog will link to it from their own weblog. Various methods exist to find out about the inbound links to one's own weblog. At the time of writing, the most useful are referer listings and *weblog ecosystem* data. Referer listings are provided either as part of Web hosting service or by a third party. They give data on what particular pages linking into a weblog were used to reach it. Weblog ecosystem data is gathered by robots which exclusively surf weblogs and keep track of the links that connect them. Technorati¹¹, Blogdex at the MIT Media Lab¹², and the Blogging Ecosystem¹³ are widely used ecosystem tools. Figure 3.7 illustrates one such tool.

One way to enhance the value of pull is to post the questions one is asking oneself to one's weblog. This often generates replies on other weblogs.

3.5.4 Regular reading

Regular reading simply consists in looking up what interesting people have to say, either by visiting their weblogs or by reading content in a personal news aggregator. Reading is valuable because it provides frequent exposure to new ideas that may support or complement the researcher's interests.

¹¹URL:<http://technorati.com>

¹²URL:<http://blogdex.media.mit.edu>

¹³URL:<http://www.myelin.co.nz/ecosystem>

Technorati

Member Sign-In

COSMOS	GOOGLE JUICE	GOOGLE RANK	WATCHLISTS	MEMBERS	HELP
INBOUND LINKS TOP 100 INTERESTING RECENT BLOGS					
TECHNORATI STATS					
23,791 blogs watched 4,136,591 links tracked					
HOT NEWS					
BlogdexTop 10 Last Updated: 04:19AM 1/17/2003 PST					
1. 'The United States of America has gone mad' (Cosmos) 2. CNN.com - U.N. Inspectors find empty warheads in Iraq - Jan. 16, 2003 (Cosmos) 3. James Lileks Fisks John 'America Has Gone Mad' LeCarre (Cosmos) 4. Lessig: Whither the Supremes' Constitutional commitment? (Cosmos) 5. The New York Times Profiles Glenn Reynolds Of Instapundit (Cosmos) 6. FrozenCPU.com :: Product = PC 12v Cigarette Lighter Adapter Kit (Cosmos) 7. sacked by his fucknut employer for running a blog (Cosmos)					
Blog URL: <input type="text" value="http://radio.weblogs.com/0110772"/>					
<input type="button" value="Get Link Cosmos"/> <input type="button" value="Make this a Watchlist"/>					
166 Inbound Blogs, 360 Inbound Links to Seb's Open Research Blog last updated 3 days 22 minutes ago					
Ranked by freshness (Rank by Blog authority)					
1: Kushal's Blog 1 inbound blog, 1 inbound link (Last updated 10 hours 59 minutes ago) n passing , answer bus , the lawsuits against reality tv . Some of these were found in Geellog , which was kind enough to link to me. Speaking of blogs, the searches for "marijuana the sims" have tapered off. And I continue to enjoy the meta discussions on Seb's Open Research . (Link created 11 hours 4 minutes ago) (Cosmos)					
2: Seblogging 54 inbound blogs, 251 inbound links (Last updated 3 hours 12 minutes ago) Weblogs and blogging for knowledge management via Seb Quote: "Blogging is a great enabler of knowledge sharing. It plays on people's natural inclination to share their thoughts and help others - the inverse of the widely accepted myth that people are reluctant to share knowledge. Comment: Short article about (Link created 16 hours 1 minute ago) (Cosmos)					
3: One Pilgrim's Walk 1 inbound blog, 1 inbound link (Last updated 16 hours 7 minutes ago) Sébastien Paquet of Université de Montréal (I'm not sure whether student or faculty -- but I'm not sure it matters in this case) has a "paper" online: Personal knowledge publishing and its uses in research . A nice start to the issue of what is a weblog, and what weblogs might become. Paquet offers this definition of a blog: (Link created 18 hours 45 minutes ago) (Cosmos)					
4: One Pilgrim's Walk 1 inbound blog, 1 inbound link (Last updated 16 hours 7 minutes ago) Weblogs, what they are and ? (Link created 18 hours 45 minutes ago) (Cosmos)					
5: One Pilgrim's Walk 1 inbound blog, 1 inbound link (Last updated 16 hours 7 minutes ago) Sébastien Paquet of Université de Montréal (I'm not sure whether student or faculty -- but I'm not sure it matters in this case) has a "paper" online: Personal knowledge publishing and its uses in research . A nice start to the issue of what is a weblog, and what weblogs might become. Paquet offers this definition of a blog: (Link created 18 hours 45 minutes ago) (Cosmos)					

Figure 3.7: Technorati inbound link data for Sébastien Paquet's *Open Research* weblog, captured on January 17, 2003. The words in bold characters are links to sites referencing that weblog.

Furthermore, regular reading is the most important activity for network building because it allows a researcher to strengthen and maintain relationships with other people who he would otherwise have no or little contact with. It can thus be seen as a way to strengthen so-called inter-community “weak ties” [Gra73, Rue02] and better exploit their natural tendency to generate new knowledge.

Of course, the speed with which this process can happen will depend on the time invested by all parties involved. Someone who is successful at networking using this tool, however, will not need to expend as much time and energy into other forms of networking. Chapter 6 empirically examines, among other things, the relationship between frequency of updates and connections established.

3.6 Limitations of personal knowledge publishing

Here I describe the main limitations of personal knowledge publishing in the context of interdisciplinary research communication.

Technical limitations. Three limitations fall into this category.

Bandwidth. Currently, the overwhelming majority of weblogs consist of text, sometimes with small pictures thrown in. Richer media formats would presumably enable better, new forms of communication. However, they consume a sizable amount of bandwidth. Although experiments with “video blogging” is underway, a successful video weblog network might well be too expensive to be usable at present.

Mobility. For most people, editing a weblog requires that they sit down at their computer, at home or in the office. In certain cases it might be useful to update one’s weblog remotely; the prime example is conference settings [Dys02]. This

will have to wait until wireless access to the Internet and mobile devices become ubiquitous.

Data security. Some weblog services hold their users' data in a central location. Consequently, some people may be unsure about their ability to maintain the integrity of their data. Desktop-based weblog tools exist that keep the user's data locally, in which case the user takes care of his data himself. However, he then has to carry the computer along with him wherever he travels and connect it to the Internet, which may be inconvenient.

Cognitive limitations. Three limitations fall in this category as well. The first has to do with the difficulty of locating knowledge after a period of time, the second with the fact that much of the knowledge remains in tacit form in personal publications, and the third with the difficulty of making the most important ideas stand out from the rest.

Locating knowledge. Personal knowledge publishing is free-form and lacks any reliable systematic organisation other than chronology. It is more similar to a monologue or conversation record than to a true knowledge organizing tool. As a result, it is difficult to locate particular thoughts after a while, even for the author, unless one remembers specific words from a particular post. Knowledge discovery in that context relies mostly on serendipity. Recording knowledge for easy retrieval requires going beyond this form of communication.

Important content does not stand out. By default, the only content that is given privileged status in personal knowledge publishing is the fresh content that appears on the front page. Nothing distinguishes "classic pieces", older content that is especially insightful, from more ordinary commentary. This situation is especially problematic when one wants to have a picture of a newly encountered author

that shows what he or she is able to achieve, in order to decide whether to return to his or her site.

Tacit knowledge. Knowledge as represented in personal knowledge publishing relies heavily on assumptions that are not made explicit. The author explains things as he sees them and typically glosses over many details that may not be understood by someone who has just entered his flow of thought. This results in a cultural barrier which makes the knowledge contained in a personal knowledge publication less accessible to people not acquainted with the author's background knowledge. Obviously this is a characteristic of any mode of informal communication. Expository style varies among individuals, and some will be successful at making their ideas more widely accessible, but as long as informality is present there will remain ambiguity left for readers to resolve by guessing.

The three problems above are indications that, although personal knowledge publishing is well-suited for interdisciplinary networking and for enhancing knowledge flows, it does not solve the whole problem of interdisciplinary communication, especially as regards knowledge reuse.

3.7 Suitability for interdisciplinary work

The previous sections articulated personal knowledge publishing's modus operandi in a general way. Here I describe five distinct advantages of personal knowledge publishing over other means of collaboration for work of an interdisciplinary nature.

Low cost. When envisioning interdisciplinary communication, as compared to intradisciplinary collaborations, it is often difficult to reliably assess in advance whether mutual understanding will be attained or if participants will talk past one

another. It is thus desirable to have a means of exploring common interests that is not costly. Personal knowledge publishing provides such a means. Moreover, each participant invests according to his own goals, so there are no frictions due to excessive coupling between participants.

Collaboration across time. Interdisciplinary problems often appear as ambitious and it is often unclear that one has access to the necessary human resources to solve them. Personal knowledge publishing enables people to collaborate with people whom they don't yet know, by giving away some of their knowledge in advance of knowing who might be interested in using it. This makes it possible to tackle ambitious problems even when the resources don't seem to be available in one's initial network, because new people can be expected to eventually find their way to one's weblog and bring needed parts of the puzzle.

Personal ground. Traditional means of research communication have some sort of link to a particular parent discipline. By contrast, a personal knowledge publication reflects the interests of one individual and is thus not necessarily bound to any discipline. This provides an opportunity for self-expression that is conducive to interdisciplinary work, because there is no well-defined "target audience" to whose tastes one has to cater to. The researcher writes entirely on his own terms, without any obligation to use any particular specialised language. This particularity has been called *semantic autonomy* in the context of distributed knowledge management theory [BCMN02].

Fine-grained idea diffusion. Interdisciplinary ideas are often quite difficult to communicate and thus spread with difficulty even when they are useful. They are best transmitted in one-to-one fashion, between two people who trust one another enough to actually expend the effort required to wrap their brains around outlandish ideas. People are more willing to consider new ideas when they come from trusted sources. Because of this, the links that result from personal knowledge

publishing provide a channel that is appropriate for diffusion of such ideas.

Unlimited space. Traditional communication structures, such as journals and conference, have space and time constraints. Because of this, more audacious ideas have a more difficult time making it to the public eye. In personal knowledge publishing there are no such restrictions, which makes it easier to experiment with wilder ideas. In the worst case, the ideas remain obscure as no other work builds on them, but their presence does not threaten the rest of the space.

These considerations lead me to formulate the following two theses, the validity of which will be evaluated in Chapter 6:

1. Personal knowledge publishing, as a practice, allows researchers to establish meaningful personal contacts with researchers in fields other than their own.
2. The medium of personal knowledge publications allows knowledge to be shared across fields.

Chapter 4

Open Shared Knowledge Repositories

The limits of our language mean the limits of our world.

Ludwig Wittgenstein

In this chapter, I introduce a second tool for knowledge sharing that I call the *open shared knowledge repository*, and I explain how it can be used as a way of fostering interdisciplinary communication. In a nutshell, this tool enables communities of interest to emerge and structure themselves outside the scope of any particular existing discipline (though borrowing ideas from many disciplines is obviously to be expected).

The concept of an open shared knowledge repository synthesises observations of how people share and together create knowledge in face-to-face situations, and in Web-enabled collaborative systems. Although there are many such systems in existence, there is a dearth of literature on the subject and many basic concepts have not been understood

and described explicitly. Hence, one of the goals of this chapter is to provide a foundation for discussing these systems.

It will be seen that an open shared knowledge repository can usefully complement personal knowledge publishing by addressing several of the limitations that I have identified in the previous chapter.

Section 4.1 highlights the need for researchers interested in interdisciplinary problems to build a “knowledge commons” around such problems, providing a specific motivation for the ideas in this chapter. Section 4.2 gives a definition of an open shared knowledge repository, and Section 4.3 provides an illustrative example of such a repository. Section 4.4 discusses implementations of open shared knowledge repositories. Section 4.5 explains how open shared knowledge repositories work to allow knowledge from various disciplines to be combined in useful ways. Section 4.6 identifies limitations of open shared knowledge repositories with respect to interdisciplinary knowledge sharing. Finally, Section 4.7 explains why open shared knowledge are especially advantageous for interdisciplinary knowledge sharing as compared to alternatives.

For the sake of brevity I will often use the shorter term “knowledge repository” or simply “repository” to refer to an open shared knowledge repository.

4.1 The need for open collaboration

Let us go back to Jim, the researcher from the introduction, to motivate the ideas of this chapter. Suppose that Jim has begun networking with a few other researchers who are also interested in the general problem of “exploiting the struc-

ture present in man-made geometries in numerical computing problems". The problem has a little to do with computer graphics, with numerical computing, with architecture, and perhaps also with pattern recognition, but it does not fit squarely within any of those fields.

Together, through conversation, Jim and his newfound collaborators have managed to agree on a name for the topic which constitutes their common interest, have begun fleshing out the most relevant concepts and promising ideas, and have exchanged references to relevant literature.

Now, they believe the issue is important, interesting, and big enough that surely many other people, unbeknownst to them and perhaps lying outside their combined social network, have an interest in that issue (either focal or peripheral). One thing Jim and his colleagues want to do is to *bring more people* into the discussion in order to enrich it and enable knowledge to grow faster. Another thing they want to do is to *organise the knowledge* that their discussions generate in such a way as to make it usable and accessible, both by themselves and by outsiders, and to *develop a common vocabulary* to make communication easier among themselves.

The challenge before them is to find an appropriate way to do these things.

To be fair, I must acknowledge that this "open group" orientation may not be the knee-jerk strategy of all researchers who find themselves in a similar situation. Career-wise, keeping things under wraps until something publishable comes out might be a better strategy, especially if a problem has been identified that seems to be tractable with the resources of the small team. Obviously, however, when a group chooses secrecy over openness it cannot reap the benefits of openness, among which one must notably count the ability to solve more ambitious problems.

In this work I assume the open approach, but a closed one could also be adopted

by restricting access to the shared knowledge. An approach that is restricted to a closed circle of participants would yield the same benefits as the open approach, except that it would not enable further aggregation and would thus limit the growth of shared knowledge to what the members of the closed team can produce.

In full generality, the problem is this. Suppose that a number of people in different disciplines have found a common interest in an emerging theme that they believe warrants further exploration, and that through conversation they have identified a few key ideas and relevant documents that they want to expand upon. How can they grow a community and a structured body of knowledge around those core elements?

Such an activity has been referred to by Bonifacio *et al* [BCMNO2] as the *federation of knowledge nodes or KNs*:

“KNs should be able to spontaneously federate creating groupings and communities of nodes that display a common interest. Such an interest could be given by the goal of maximizing the opportunity to be found by other KNs (being part of a visible group), by the need to certify the type or the quality of a knowledge (through the filtering of members), or the issue to protect content and secure knowledge access from unauthorised KNs (through access policies). K-Federations can simplify interaction processes because a request can be sent to a group rather than to individual KN, decreasing the number of interactions, or because knowledge retrieved from a KN that belongs to a group has presumably a certified quality.”

What I propose in this chapter is to use open shared knowledge repositories as a tool for achieving that goal.

4.2 Defining the term

In this section, I describe the structure of open shared knowledge repositories. This concept is directly inspired from the idea of a *wiki*, which I briefly describe here. (“Wiki wiki” means “quick” in the Hawaiian language, and is the origin for the name.)

The wiki idea is due to Ward Cunningham[LC01]. Cunningham’s prototype implementation¹ has given birth to many variants, so that a precise definition of wiki does not exist, but the general consensus is that a wiki is *a collective website where any participant is allowed to modify any page or create a new page*. In the most open implementations, any visitor may edit any page, without even logging in. Revision histories (including author/machine identification) are accessible, making it easy to restore an earlier version of a page to correct erroneous information or to undo a deletion.

Typically, a simplified HTML syntax enables easy formatting and facilitates hyperlinking within a wiki. For instance, CamelCase words (capitalized words strung together) and expressions inside double brackets will automatically translate into links to other pages within the wiki. *Backlinks* enable visitors to obtain a list of the pages that link into any particular page. A continuously updated list of the recently changed pages lets them locate activity in the wiki. Participants may obtain a username, which will be attached to the changes they make when they are logged in.

Wikis took off as a new way to conduct discussions and collaboratively edit documents via the World Wide Web. In recent years, several different wiki systems have been set up, some public, some private, and have been put to several different uses. One of the most high-profile wiki projects is Wikipedia², which aims to

¹Still alive at the URL <http://c2.com/cgi/wiki>.

²URL: <http://wikipedia.org>

create a free encyclopedia, and has reached the respectable size of 100,000 articles a little while after its second anniversary in January 2003. Another example of a successful wiki is the Swedish general reference site susning.nu³[Aro02].

Figure 4.1 shows a page in the Wikipedia, for purposes of illustration.

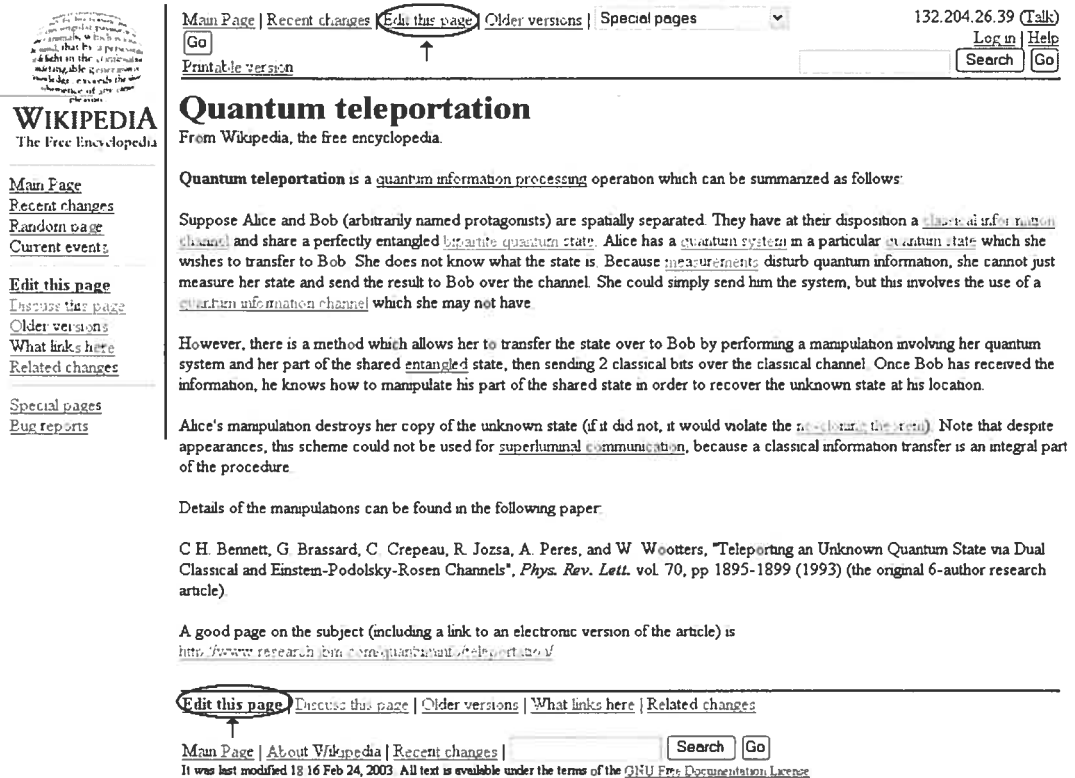


Figure 4.1: The “Quantum teleportation” page from the free encyclopedia project Wikipedia. The “Edit this page” link (circled) enables visitors to modify the page’s contents.

My observations of real-life wikis were the starting point for the definition of an open shared knowledge repository that I provide in this chapter. More details on the technical and social aspects of implementing such resources are found in the following sections.

Just as was the case for personal knowledge publishing, the definition of such

³URL: <http://susning.nu/>

an open shared knowledge repository involves two components, one social and the other technological. Knowledge repositories cannot be understood without referring to both aspects.

From the technological standpoint, an open shared knowledge repository follows the simple basic structural model of the World Wide Web: it is made up of *pages*, or documents, each with a title, and links between pages. Pages' contents consist of a descriptive title, plus text, code, images, links to other pages in the repository, and links to other, external documents. The benefits of hypertext (such as the possibility of self-directed exploration) are thus inherited.

From the social standpoint, open shared knowledge repositories have five defining features:

- 1) *A knowledge repository is a collective artifact.* It acts as a beacon for a community of people. The content of the knowledge repository is wholly accessible to all members of the community; it is managed by the community, for the benefit of its members. This is the reason for using the adjective "shared".
- 2) *The community is open.* Anyone with the technological means to access the repository can contribute to it, even if they do not know a single person in the community a priori.
- 3) *The content has shared meaning.* The content of the knowledge repository is meant to represent the community's agreed-upon knowledge.
- 4) *The content is freely accessible* and may be freely reproduced or redistributed.
- 5) *Authors can make themselves known.* Although anonymous contributions may be accepted, participants have the option of registering their name so that it will be attached to their contributions (although authentication is not necessarily implemented).

Whereas the WWW is a free-for-all in terms of structure, an open shared knowledge repository is more specifically constrained. Each page has a descriptive title, which tells how the community refers to the particular theme or concept that is discussed in the page, and which can be used in other pages to create links to that concept. The content of a page should reflect what the community members agree is a good description of what they collectively know about the topic. It should also point to the pages describing topics that relate to that topic.

4.3 An illustrative example

Here I illustrate the notion of an open shared knowledge repository using the real-life example of a site called Meatball Wiki, which is concerned with online communities.

Figure 4.2 shows a snapshot from a particular page in the Meatball Wiki, taken on December 12, 2002.

The title of this page is “HyperMedium”. The text that is labeled with number 3 is the contents of the page proper. The number 3 refers to the third criterion in our definition above: this piece of text is taken to specify the meaning of the term “HyperMedium” within the community. Other hyperlinks are apparent on the page. On Meatball Wiki, similarly to other wiki systems, page titles are written in so-called *CamelCase* - strings of words with each first letter capitalised. This makes linking easier: any CamelCase word is automatically turned into a link into another repository page with that word as title. (If the page does not exist, a link with a different appearance is made to indicate the fact.)

Figure 4.3 shows the recent revision history of the same page. It can be seen that two different authors have recently modified the page, leaving a signature (crite-

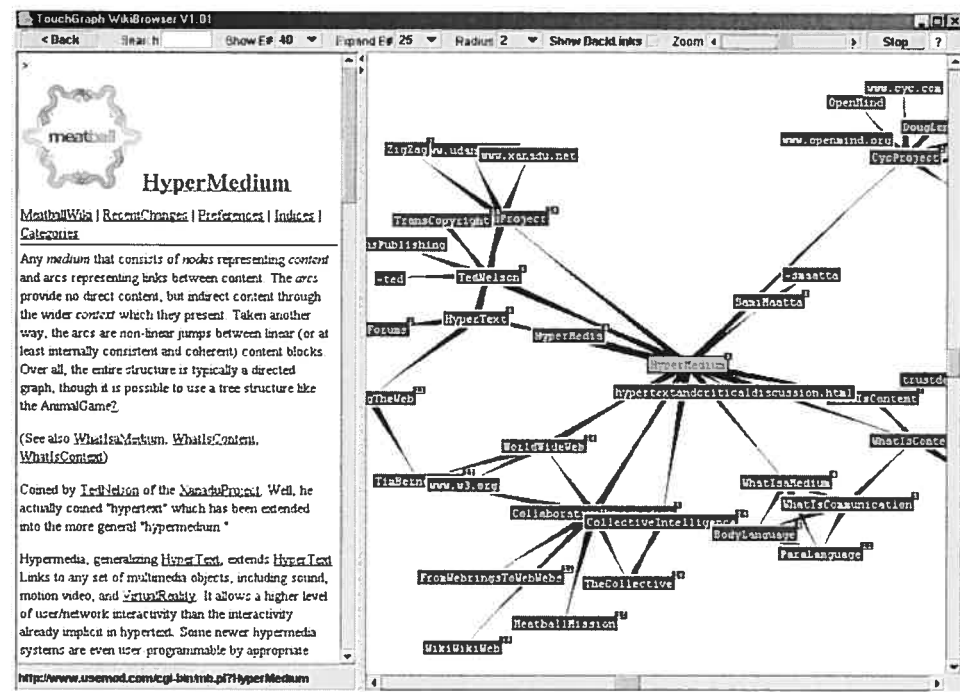


Figure 4.2: The page titled “HyperMedium” in Meatball Wiki, and a visualisation of its relationship to neighboring pages.

riterion 5), which is evidence of the repository being a collective artifact (criterion 1). Criteria 2 and 4 cannot be readily illustrated because they are implicit in the usage conventions of the resource by the community.

Figure 4.4 shows the results of a search of the Meatball Wiki site for links to the “HyperMedium” concept (these are called the *backlinks* of the page). It indicates that the concept has been linked to other concepts and gives a certain amount of evidence that the community relies on this concept. Though it cannot in itself be proof that the meaning of “HyperMedium” is universally agreed among community members, one can expect that as the reuse of a concept happens in more places, any disagreements as to its meaning will become apparent to members, who will try to resolve them if they care about the concept.

History of HyperMedium

[MeatballWiki](#) | [RecentChanges](#) | [Preferences](#) | [Indices](#) | [Categories](#)

☐ ☒ [Revision 14](#) . November 26, 2002 3:34 pm by [SuratShah](#) [cf [AboutLinks](#)]
☒ ☐ [Revision 13](#) . (edit) October 25, 2002 10:24 am by [DavidAndel](#) [+ cat]

[Compare](#)

Difference (from prior major revision) (no other diffs)

Changed: 9c9

An excellent practical example is the [WorldWideWeb](#) which demonstrates linear *pages* for the nodes, and *hyperlinks* as the arcs. Another excellent example is a dictionary. The nodes are the definitions, the arcs are the words used in the definitions (they refer to other dictionary entries). Another example is a "Choose your own adventure" book where the links are page references at the bottom of each mini-chapter. In this latter case, note how the links are out-of-line (not directly embedded in the content) versus in-line (directly embedded in the content).

An excellent practical example is the [WorldWideWeb](#) which demonstrates linear *pages* for the nodes, and *hyperlinks* as the arcs. Another excellent example is a dictionary. The nodes are the definitions, the arcs are the words used in the definitions (they refer to other dictionary entries). Another example is a "Choose your own adventure" book where the links are page references at the bottom of each mini-chapter. In this latter case, note how the links are out-of-line (not directly embedded in the content) versus in-line (directly embedded in the content). See [AboutLinks](#) for a taxonomy of links.



1. 5

Figure 4.3: The revision history for the page titled "HyperMedium" in Meatball Wiki, along with the differences between the last two versions.

Backlinks for: HyperMedium

[MeatballWiki](#) | [RecentChanges](#) | [Preferences](#) | [Indices](#) | [Categories](#)

8 pages found:

[CollaborativeHypermedium](#)
[HyperMedia](#)
[HyperMedium](#)
[LinkGraphView](#)
[TGWikiBrowserScreenShots](#)
[TedNelson](#)
[TouchGraphWikiBrowser](#)
[WorldWideWeb](#)



Figure 4.4: The backlinks for the "HyperMedium" page in Meatball wiki. (Note that it links to itself.)

4.4 Implementation of open shared knowledge repositories

In this section, I describe two practical ways of implementing an open shared knowledge repository on top of the World Wide Web infrastructure.

4.4.1 Traditional Wiki

The current architecture of the Web and the available scripting languages make it relatively easy to implement a wiki. Several implementations of the wiki concept exist and are available on the Web⁴

As regards the social aspects, since a wiki is generally accessible via the Web, it is *de facto* a collective artifact. The Web protocols (TCP/IP, HTTP) are open to all by default, which makes the community open as well. The following two social facets of the definition (shared meaning and free access), are a matter of policy, to be agreed among the wiki's users. The last social facet, attribution of changes, is made possible (albeit without authentication in most implementations) by the mechanism described above.

4.4.2 Individual Custody Wiki

What I call an “individual custody wiki” consists in a system that presents the same features as a traditional wiki, with the following exception: each individual page is “owned” by a particular user, who is the only one allowed to modify the page. Recommendations by other users may be appended, however. A user may

⁴See, e.g., the list at the URL <http://www.usemod.com/cgi-bin/mb.pl?WikiClone>.

transfer ownership to another when he is no longer interested. A user who doesn't respond to recommendations by other users on a particular page loses ownership after a certain time.

4.5 Theoretical considerations: how open shared knowledge repositories work

Here I develop a theoretical rationale for the ability of the tool just described to fulfill a need for simultaneously building a community and collaboratively structuring and synthesizing a body of knowledge.

I first explain how an open shared knowledge repository enables knowledge to be represented and found, which is its most obvious use. Then I explain the three roles that the body of knowledge plays in building a community and helping generate additional knowledge. First, it acts as an *attractor* that pulls new participants into the community. Second, it acts as a *glue* that holds the community together; Third, it furnishes a *bedrock* upon which new knowledge can be created.

At the end of the section, I explain why open shared knowledge repositories offer a medium that is especially well-suited to interdisciplinary communication, as compared to alternative ways of achieving the same goals.

4.5.1 Using the repository to represent and find knowledge

To the casual observer, an open shared knowledge repository looks just like an ordinary Web site. It can thus be explored and searched just like one would search the World Wide Web itself.

However, its specific nature makes search exploration easier in several ways. The *first advantage* is that of relevance: restricting text searches to the repository guarantees that one will find discourse that relates to the overall theme of the repository. The problem of irrelevant results due to words that have several meanings is reduced.

A *second advantage* results from the uniform encyclopedia-like structure of the repository. Upon encountering a link, one can strongly expect that at the other end one will find a description of what is meant by the linked term.

A *third advantage* is the possibility to look up who made what contribution. This enables the user to easily identify people who are knowledgeable on very specific issues. (One has too keep in mind that the information might not be valid in the absence of authenti

A last advantage is that visitors can very easily contribute new content, either directly into relevant pages, or by creating new pages. Moreover, any visitor who finds relationships between topics that have not yet been made explicit in the repository through linking can easily let others benefit from his discovery, simply by contributing new links into the appropriate pages. Thus as the site is used, it becomes more organised. Obviously, the quality of that structure depends on the amount of effort put into this activity by participants.

4.5.2 The body of knowledge as an attractor

Any intellectual community cannot grow unless it makes itself visible and at least minimally understandable to outsiders.

The aggregation potential of a community mainly depends on three factors which I will explain here: *visibility*, *user-friendliness*, and *perception of continued rele-*

vance to visitors.

Visibility means that there exists a community that is larger than the group of participants, and that the people in this greater community can be exposed to the repository in one way or another.

The means of visibility depend on the implementation medium of the repository. In the case of a physical bulletin board, if it were to be located in a public place where people come and go and can see the board, then this would provide a means of visibility to the community of people who pass in that place.

In the case of HTTP-based implementations, the most obvious way to obtain visibility is to make the repository pages available on the World Wide Web. Visibility will come as a byproduct of inbound hyperlinks, appearance in search engine results, and possibly also through publicised URLs.

User-friendliness. Outsiders will not stick around for long if they do not easily grasp what the community has to offer to them. User-friendliness reflects whether the repository is understandable and appealing to first-timers.

Continued relevance. Visitors must not only find value in the content, but also see that the content evolves in time. This is the first step to establishing a continued relationship with the resource.

Given visibility, user-friendliness, and continued relevance, some people will end up somewhere in the knowledge repository, understand at least rudimentarily what it is and what it gives to them, perceive value in it, understand that the repository is continually updated, and elect to return at another point in time, e.g. by remembering the URL or by bookmarking or blogging the site. This process may be called the attraction effect of the knowledge repository.

4.5.3 The body of knowledge as community glue

The body of knowledge also plays an important role in holding the participants together. It becomes a shared resource whose integrity participants care about. Below we explain the close link that exists between the sense of community and the issue of content quality control.

An open shared knowledge repository could be likened to a dynamic community encyclopedia. However, traditional encyclopedias have editors-in-chief, who are in a position of authority and who, as such, grant or deny others the opportunity to contribute to the collective work. This allows them to shape the result of the overall effort, and usually ensures that the resulting knowledge is as reliable as they are. What happens when everyone in the (open) community is an editor, in the absence of centralised control?

The answer depends on the nature of the interactions between the community members. In the absence of authority, disagreements about the contents of a particular page have to be resolved through discussion. If they are not, the page in question will constantly be modified until the conflict is resolved or abandoned.

Cohesion, the ability to reach consensus within the group, determines the amount of conflict that will be observed. Obviously, all other things being equal, a community with a high degree of cohesion will grow its ranks and its knowledge repository faster than one that doesn't have it. It is obviously quite possible that in some cases there will not be enough cohesion (or commitment to achieve it) to ensure a sustained growth; then the repository will stop evolving in any meaningful way. However, it should be kept in mind that too little diversity is likely to result in intellectual stagnation. For the growth of knowledge it is thus necessary to remain between the extremes of chaos and uniformity.

A commitment to achieving cohesion can arise when the participants believe that

they will benefit from building an intellectual common ground more than if they did not collaborate. Possible benefits include, but are not limited to: obtaining knowledge that they could not obtain elsewhere, and establishing new personal connections with community members.

And this is how the body of knowledge acts as a kind of community glue. The community is permeable: it allows individuals to come and go as they please. The people who have a stake in the integrity of the body of knowledge will stick around and participate in maintaining and improving it, and those who don't will be driven away.

Or will they? Up to now we have not considered explicitly the possibilities of such antisocial behaviours as vandalism or spamming. People who engage in such behaviour are not motivated by knowledge sharing, but rather by pushing particular information of their choosing into eyeballs.

Insofar as the integrity of the shared knowledge is concerned, vandalism and spamming do not constitute special cases to be distinguished from, say, well-intentioned but erroneous statements. Any content that is not agreed upon is filtered by the community. A good implementation of an open shared knowledge repository makes this trivial to do by the first well-intentioned person who recognises the offending act and fixes the situation. The willingness of members to "keep the place tidy" and the ease to do so drives down the incentive for such acts. Repeat offenses are not ruled out, but in a *cohesive* community the work of preserving integrity is divided among several (though presumably not all) of its members, which makes it resilient to such attacks.

A word here is in order regarding property no. 4 of our definition, "The content is freely accessible and may be freely reproduced or redistributed." This property is important to provide assurance to participants that their work will not be hijacked and its diffusion restricted through copyright enforcement or other means. Many

people want their ideas (or reputation) to spread as much as possible and will not contribute to a knowledge sharing effort without this assurance. (This universal accessibility is actually one of the fundamental principles of science.)

4.5.4 The body of knowledge as foundation for new knowledge

The requirement that each page in a knowledge repository describe a topic or concept and give a *name* to that concept gives the participants the power to go beyond just collecting and linking existing pieces of knowledge. It offers them an easy way to construct arbitrarily elaborate new ideas within the repository.

This is because names are extremely useful as shorthand representations of complex ideas. Any technical term readily illustrates this fact. Of course the power of names or symbols is a general property of any extensible human language, but the hyperlinked structure of open shared knowledge repositories exploits it in a natural and convenient way.

To build upon already existing concepts in the repository, one simply articulates new ideas by referring to these other concepts by name (and linking to them). People who later wish to understand the new idea can very easily look up the relevant concepts for which they need a refresher or a primer.

In this way, new meanings and accompanying names are created, and the community's knowledge grows spontaneously around hotbeds of activity.

4.6 Limitations of open shared knowledge repositories

Here I discuss the limitations of open shared knowledge repositories in the context of interdisciplinary research collaboration.

Technical factors. Three main limitations are of a technical nature. First, *centralisation* is an issue. All current implementations of open shared knowledge repositories hold data in a centralised location, which means that scaling the number of participants can become costly to the hosting provider. A more decentralised model, similar to the World Wide Web itself, would be more desirable in order to accomodate large numbers of participants (though it is unclear that a creative interdisciplinary community can be expected to scale to large proportions). This could be done using mirror sites and database coordination protocols, and by decentralizing such functions as search.

Second, *vulnerability to robot attacks* is an issue. Open shared knowledge repositories assume that actual humans are interfacing with it through the Web form they provide. However, nothing prevents a malicious person from programming a robot to submit data very rapidly to corrupt the knowledge repository, possibly faster than any human (team) can repair it. Various workarounds can be devised⁵, but ensuring complete protection against a determined attacker is probably quite difficult. Fortunately, open shared knowledge repositories are unlikely to make enemies, and the challenge factor in corrupting an open repository is so low as to make it an uninteresting target.

Third, *rich media* is not very well supported by current implementations. Being

⁵See e.g. von Ahn, Blum, Hopper, and Langford's "Completely Automated Public Turing Test to Tell Computers and Humans Apart" (CAPTCHA) project; URL: <http://www.captcha.net/>

able to include and hyperlink to, into, and from audio and video files could certainly be a helpful complement to regular text, both in terms of making the user experience more enjoyable and facilitating learning.

Psychological/social factors. The first limitation has to do with the issue of *individual credit*, especially as it relates to researchers' career issues. The second has to do with ensuring the content has *shared meaning* in the community, and the third relates to the long-term *accessibility* of the material to outsiders.

Individual credit. In the current context of research, it is necessary for researchers to demonstrate their individual contributions to the advancement of knowledge. This can be difficult to do for work put in an open shared knowledge repository because of the inherent collaboration dynamics.

Individual pages in an open shared knowledge repository may be created by one person and then modified by several other people. Even given the revision history of a page, after a while it becomes too difficult to disentangle the respective contributions of the editors who have been involved in evolving the page.

This difficulty in precisely attributing credit can limit participation, especially from those researchers whose professional survival hinges on showing what their specific contributions are.

In a way this problem is none other than that encountered by the members of a research team who work intimately enough together that keeping track of "who did what" becomes impractical due to the constant bounce of ideas. In this light, an open shared knowledge repository is the workspace for a worldwide, extensible team, and credit for discoveries, inventions, or insights should perhaps be attributed to the whole team. However, because the team is open, the derivation of individual credit remains problematic.

It must be pointed out, however, that free and spontaneous exchange of ideas and intensive collaboration is the best way to advance knowledge. Indeed, some scientific undertakings, such as the high-energy physics experiments and genomics projects of the last decade, could not have happened without large-scale collaboration – and the corresponding rise in number of co-authors on articles.

Shared meaning. The issue I wish to raise here is that of ensuring that everyone in the community uses the terms in the knowledge repository in the same sense. This is important because misunderstandings are bound to occur if such is not the case. If meanings drift, confusion will set in and it will become increasingly difficult to build upon the ideas present in the repository.

As it turns out, an open shared knowledge repository does not provide any guarantee that meaning is shared reliably across the community. This limitation is however common to all means of communication that use informal natural language. The next chapter puts forth a possible means of overcoming this limitation.

Accessibility to outsiders. As an open shared knowledge repository grows, the regular contributors start building more and more on top of each others' newly defined concepts. However, in this process, outsiders are progressively left behind as a new language is effectively being developed, and language barriers appear around it. Even if they have considerable related knowledge, outsiders don't have shortcuts for understanding the material. This poses an obstacle to knowledge flow in both directions across the community boundary and eventually hinders the community's growth.

4.7 Suitability for interdisciplinary work

The previous sections articulated the open shared knowledge repository's *modus operandi* in a general way. Here I describe five distinct advantages of a Web-based repository for work of an interdisciplinary nature over other means of collaboration.

Low cost. When envisioning interdisciplinary collaborations, as compared to intradisciplinary collaborations, there are generally more unknowns at the outset and it is often more difficult to reliably assess in advance whether success is likely or not. It is thus useful to have a means of exploring common interests that does not ask of participants that they risk investing considerable time and/or funds into what may turn out to be a dead-end. An open shared knowledge repository takes little time and money to set up, so it provides such a means.

Speed. Much interdisciplinary work is not even attempted because the time it would take to gather relevant knowledge and people to reach a sufficient stage of development is too considerable. By contrast to other means of collaboration, an open shared knowledge repository enables thoughts to be shared publicly and reviewed instantly. This makes it possible to undertake projects that would be too ambitious for other means of collaboration.

Collaboration with strangers. A knowledge repository that is open enables people to collaborate with people whom they don't yet know. This makes it possible to tackle ambitious problems even when the resources don't seem to be available, because it is always possible that new people will find their way to the repository and bring needed parts of the puzzle.

Neutral meeting ground. Laboratories in research departments and various collaboration spaces that are attached to a particular journal or conference are explicitly

linked to a parent discipline. This can prove intimidating for collaboration with people in other disciplines because of the inherent asymmetry between the parent discipline and the other ones. An open shared knowledge repository does not need to depend on any one preexisting disciplinary structure. It can thus provide a common space that is not primarily attached to a particular discipline. This helps all participants feel comfortable by placing them on equal footing; no one is put in a position of “authority”.

Unlimited space. Traditional communication structures, such as journals and conferences, have space and time constraints. Because of this, more audacious ideas have a more difficult time making it to the public eye. In an open shared knowledge repository, there are no such restrictions, which makes it easier to experiment with wilder ideas. In the worst case, the ideas remain obscure as no other ideas build on them, but their presence does not have the effect of hiding the other ideas present in the repository.

These considerations lead me to formulate the following two theses, the validity of which will be evaluated in Chapter 6:

1. Open shared knowledge repositories allow researchers to build a research community around an interdisciplinary theme.
2. Open shared knowledge repositories allow researchers to build a body of knowledge around an interdisciplinary theme.

Chapter 5

Navigable Synthesis Ontologies

*Analogy pervades all our thinking, our everyday speech
and our trivial conclusions as well as artistic ways of expression
and the highest scientific achievements.*

George Pólya

In this chapter, I introduce a third and last tool for interdisciplinary knowledge sharing: the *navigable synthesis ontology*, which extends the open shared knowledge repository by giving it a well-defined and uniform structure. In a nutshell, this tool facilitates the reuse of knowledge across fields by enabling cross-disciplinary analogies to be organised in a coherent manner. It will be seen that a navigable synthesis ontology can usefully complement personal knowledge publishing and open shared knowledge repositories by addressing several of the limitations that I have identified previously.

Section 5.1 highlights the need for a systematic organisation of knowledge that enables bridges to be built across disciplines. Section 5.2 gives a definition of a navigable synthesis ontology, and Section 5.3

provides an illustrative example of such a system. Section 5.4 discusses implementations of navigable synthesis ontologies. Section 5.5 explains how navigable synthesis ontologies work to enable knowledge to cross disciplinary boundaries. Section 5.6 identifies limitations of navigable synthesis ontologies with respect to the objective they attend to. Finally, Section 5.7 points out why this tool should be especially useful for interdisciplinary knowledge sharing.

5.1 The need for navigable synthesis ontologies

Let us one last time return to our representative researcher, Jim. In chapter 3, I have argued that the task of accessing knowledge outside his own discipline could be made easier if Jim weaved himself into a personal knowledge publishing network. However, I have also pointed out the limitations associated with such a resource. First, knowledge discovery in that context relies mostly on serendipity and depends on Jim's intuition and patience. To succeed in finding the right people, and from there, the right knowledge, requires skill at dealing with tacit knowledge and assessing the credibility of sources.

Also, once the right person is found, she may not have the time to answer a query. Even if she is willing to answer, the delay between the query and the response might be substantial. It may also happen that the question Jim is asking himself is actually a frequently asked question in a particular field. In such a case Jim's question could be rather bothersome. From an efficiency standpoint, it makes more sense to properly document frequently asked questions that are very well understood and considered elementary in a discipline, so that people can find the answers themselves.

This is actually one of the functions of textbooks. But textbooks have a major

disadvantage: their “quantum of contribution” is very large, meaning that they are mostly an all-or-nothing proposition: it is not possible to make small contributions in this format. As a result, even material that is very well understood in some fields can take a decade or so before getting documented in a form that is suitable for consumption by outsiders.

In chapter 4, I have explained how open shared knowledge repositories enable a body of knowledge to be structured collaboratively. I have however pointed out that successful repositories run the risk of creating a new language and erecting new language barriers. What we ideally want is to make knowledge accessible to outsiders.

I now describe a number of desirable features for an interdisciplinary knowledge resource. Imagine a knowledge resource that, in addition to meeting the constraints specified in Section 1.1.6, exhibits the following characteristics:

- 1) It features only explicit knowledge, making it self-contained;
- 2) It connects analogous pieces of knowledge regardless of their belonging to different fields, in such a way that finding knowledge that relates to a specific line of inquiry requires a minimal amount of guesswork;
- 3) It can be explored autonomously by individuals;
- 4) There exists a practical means to build it collaboratively.

In the rest of this chapter I will explain the notion of a *navigable synthesis ontology* and argue that such a system would fit the bill.

5.2 Defining the term

A navigable synthesis ontology is an open shared knowledge repository that additionally conforms to the following specifications:

- 1) Each page gives one or more complete formal definitions of a concept, referencing other concept pages as needed. Definitions are complete in the sense that they specify both necessary and sufficient conditions for membership in the class corresponding to the concept. They are formal in the sense that they are rigorous and unambiguous. However, just like mathematical textbook definitions, they are written in natural, precise, but not formalized language.
- 2) Each page features references to generalisations and specialisations of the concept, that is, concepts with more inclusive and more restrictive definitions, respectively.
- 3) Each page may point to a subpage that lists things that can be considered as instances of the concept.
- 4) Each page may point to a subpage with links to external documents that specifically discuss the concept under consideration.
- 5) Each page may point to a subpage for discussion of its contents.

The “synthesis” aspect in the denomination of this object arises from its ability to link related concepts from different areas together by finding unifying concepts that capture the commonality, as is explained in more detail in Section 5.5.

By contrast to the two previous tools that were presented in Chapters 3 and 4, a navigable synthesis ontology affords a less direct form of communication between participants. It constitutes a much more structured (and thus less expressive)

intermediary between participants. The core function of the ontology is not direct communication (although it is supported, as explained in Section 5.5.2), but rather knowledge organisation for easy retrieval, in particular across disciplinary boundaries.

5.3 A short example

To help make things clearer, I have built a sample ontology that illustrates the notion defined above. I will only present two pages here; the full ontology is accessible via the Web¹ and is the subject of Section 6.4.

As specified above, all the pages in the ontology have the same structure. Figure 5.1 shows a page that precisely specifies the concept of an ordered tree.

The page begins with a list of equivalent definitions of the tree concept that point to the concepts that they involve (1). Following that, a name that is commonly used for that concept is provided. Pointers to generalisations and specialisations of the concept follow (2). There are then links to two subpages respectively listing concrete instances of the tree concept for illustrative purposes (3) and external references to documents that are concerned with trees (4). There follows a link to a subpage for discussing the concept (5). For convenience, one last link with the prefix “D/” (for “domain”) refers to the domain where the concept is predominantly used.

Following an “/Instances” link will lead the user to subpages such as those shown in Figures 5.2 and 5.3, which help understand the concept.

Following an “/External documents” link will lead the user to a subpage such

¹URL: <http://www2.iro.umontreal.ca/~paquetse/cgi-bin/uc.cgi>. See also a larger, but not openly editable ontology at <http://www.iro.umontreal.ca/~paquetse/knoweb>.

Hierarchical Ordered Rooted Tree Structure

[Unifying Concepts Wiki \(home\)](#) | [Recent Changes](#) | [Preferences](#)

A hierarchical ordered free tree structure in which a single member has no superior. That member is called the *root*.

Equivalently, a hierarchical unordered rooted tree structure in which the children of every node are ordered.

Equivalently, a hierarchical ordered free tree structure that is also a hierarchical unordered rooted tree structure.

Also known as: ordered tree (Computer Science)

Generalizations:

- hierarchical ordered free tree structure (no designated root)
- hierarchical unordered rooted tree structure (parents have unordered children)

Specializations:

- binary tree (at most two children per member)

[/Instances](#) ————— 3

[/External Documents](#) ————— 4

[/Discussion](#) ————— 5

[\(D\)/Computer Science](#)

Figure 5.1: Tree concept in a navigable synthesis ontology.

Hierarchical Ordered Rooted Tree Structure/Instances

[Unifying Concepts Wiki \(home\)](#) | [Hierarchical Ordered Rooted Tree Structure](#) | [Recent Changes](#) | [Preferences](#)

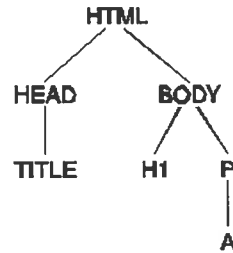
- I/Sample tree representing an arithmetical expression
- I/Sample tree representing a Web page

Figure 5.2: Available instances of the Tree concept.

I/Sample Tree Representing A Web Page

[Unifying Concepts Wiki \(home\)](#) | [I](#) | [Recent Changes](#) | [Preferences](#)

This is an instance of a [Hierarchical Ordered Rooted Tree Structure](#).



The member labeled "HTML" is the root. The root's first child is labeled "HEAD", and its second child is labeled "BODY".

Figure 5.3: One possible instance of the Tree concept.

as that shown in Figure 5.4, from where he can access documents outside the ontology that are specifically concerned with the concept.

Hierarchical Ordered Rooted Tree Structure/External Documents

[Unifying Concepts Wiki \(home\)](#) | [Hierarchical Ordered Rooted Tree Structure](#) | [Recent Changes](#) | [Preferences](#)

Course material

- http://www.cs.msstate.edu/~cs2813/fall02/section3_1.ppt: Missouri State U. Discrete Structures course - introduction to trees (PowerPoint)
- <http://www.cs.ucdavis.edu/~bai/ECS20/ecs20lecture20.ps>: UC Davis Discrete Math course - lecture on trees (PostScript)

Encyclopedias

- <http://www.nist.gov/dads/HTML/orderedtree.html>: "Ordered Tree" at the Dictionary of Algorithms and Data Structures
- <http://www.nist.gov/dads/HTML/rootedtree.html>: "Rooted Tree" at the Dictionary of Algorithms and Data Structures
- <http://mathworld.wolfram.com/OrderedTree.html>: "Ordered Tree" at MathWorld
- <http://mathworld.wolfram.com/RootedTree.html>: "Rooted Tree" at MathWorld

Figure 5.4: External documents about the Tree concept.

Following a "/Discussion" link will lead the user to a subpage such as that shown in Figure 5.5, where he can discuss the content of the page with other people who are interested in that concept.

Hierarchical Ordered Rooted Tree Structure/Discussion

[Unifying Concepts Wiki \(home\)](#) | [Hierarchical Ordered Rooted Tree Structure](#) | [Recent Changes](#) | [Preferences](#)

Aren't the words "Hierarchical" and "Tree" redundant?

No, the presence of "Tree" makes the concept more specific, as multiple hierarchies are ruled out. --Seb

Figure 5.5: Discussion subpage about the Tree concept.

Now, if we visit the page describing the more general concept of an unordered tree (where the children are not ordered) we see the content shown on Figure 5.6, which has the same structure:

Hierarchical Unordered Rooted Tree Structure

[Unifying Concepts Wiki \(home\)](#) | [Recent Changes](#) | [Preferences](#)

A Hierarchical Unordered Free Tree Structure in which there is a single member that is not the child of any other member. This member is called the *root*.

Equivalently, a Rooted Hierarchy for which each member has a single immediate superior.

Also known as: tree

Generalizations:

- hierarchical unordered free tree structure (not necessarily having a root)
- rooted hierarchy (no limit on the number of immediate superiors)

Specializations:

- hierarchical ordered rooted tree structure (ordered children - your traditional tree)

[/Instances](#)

[/External documents](#)

[/Discussion](#)

[D/Interdisciplinarity](#) [D/Computer Science](#)

Figure 5.6: Unordered tree concept in a navigable synthesis ontology.

5.4 Implementation of navigable synthesis ontologies

Implementing a navigable synthesis ontology is a simple matter of designing a specialised version of an open shared knowledge repository implementation that allows the additional specifications of Section 5.2 to be respected.

For illustrative purposes, I now describe the design of my prototype ontology system. I took as a starting point the Perl-based UseModWiki² software. UseModWiki is installed on a CGI-enabled Web server and uses its filesystem to implement a centralised database containing the data for all the pages in the Wiki, as well as maintaining change information and user records.

I modified UseModWiki so that each newly created page would already feature the predefined common structure (definition; generalisations; specialisations; instances; external links; discussion; and domains), obviously initially empty. To this end, as shown in Figure 5.7, I modified the initial content of new pages.

The instances are listed in a subpage to keep concept pages clean. I adopt the convention of prefixing the names of instance pages with the string “I/”, to clearly distinguish them from concept pages. References to material outside the ontology and discussions about the page’s contents are likewise placed in subpages.

The figures that were shown in the previous section were taken from this implementation.

²by Clifford Adams, URL: <http://usemod.com>

Editing Palindrome Word

[Unifying Concepts Wiki \(home\)](#) | [Recent Changes](#) | [Preferences](#)

'''Also known as:'''

'''Generalizations:'''

'''Specializations:'''

/Instances

[[/External documents]]

/Discussion

[[D/]]

Summary:*

☐ This change is a minor edit.

(Visit [Preferences](#) to set your user name.)

[View other revisions](#)

[Unifying Concepts Wiki \(home\)](#) | [Recent Changes](#) | [Preferences](#)

Figure 5.7: Newly created page in my implementation of a navigable synthesis ontology.

5.5 Theoretical considerations: how navigable synthesis ontologies work

In this section, I first describe how a synthesis ontology can be navigated to find knowledge that relates to a given line of inquiry, regardless of boundaries between disciplines. Then I explain how such an ontology can be built and shaped by the collective actions of many participants.

5.5.1 Using the ontology to locate knowledge

I present two main classes of knowledge retrieval problems and the associated strategies that allow an explorer to find knowledge that closely relates to a particular idea in his mind when he does not know what name (if any) has been given to that idea.

The classes are called “Analogous concept known” and “Related concept known”. I give simple examples of each case. The examples are simple ones to make the exposition unambiguous and concise.

Class 1: Analogous concept known. In the first problem, the user already knows the name of a concept that is analogous to what he is looking for. Then, assuming this concept is already in the ontology, he can take that concept as a point of departure.

Looking at the generalisations of that concept, he will select one that seems appropriate, i.e. one that abstracts out aspects that are not essential to the sought concept(s) while retaining aspects that are. Then, moving up in the hierarchy he will eventually reach a concept that only features essential aspects of what

he seeks. From that point on, he will traverse back down, selecting appropriate specialisations along the way until he finds what he is looking for, provided that it is represented in the system.

If it is not, the user may have gotten closer to what he was looking for, hopefully close enough to enable him to pursue his search using other means, for instance by following the external links in the terminal concept page.

As an illustration, we will refer to Figure 5.8. Suppose a user understands that symmetric geometric objects are objects that are invariant under a set of geometric operations. Now the user is concerned with *strings* (or *symbol sequences*) that have a similar property. Here is one way he could locate the concept he is thinking of. From the concept “Symmetric geometric object” he might abstract out geometry and generalize to “Symmetric object”, from which point he could specialise into “Symmetric symbol sequence” and again to “Reversal-invariant symbol sequence” (also known as “Palindrome”), which turns out to be what he sought.

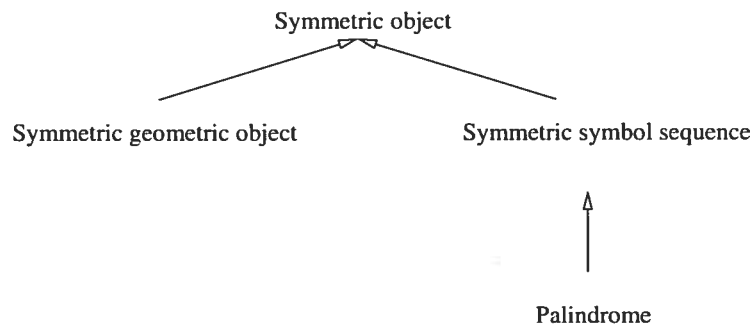


Figure 5.8: Some concepts and their interrelations in a navigable synthesis ontology. Arrows link concepts to their generalisations.

Reaching that final concept page enables the researcher to learn the precise name(s) of the precise concept he was looking for, and possibly to find authoritative documentation on that concept.

Class 2: Related concept known. In the second problem, the user knows the name of a concept that he thinks closely relates to what he is seeking, but not in a taxonomic sense. For clarity, call this concept A. In this case, that concept may serve as the point of departure for the search.

The user will then either explore one of the concepts that are involved in the definition of that A, or ask the system for the concepts that directly involve A in their own definition. (In my implementation this may be done by consulting the backlinks (Section 4.4) for the concept.)

As an illustration, suppose a researcher is looking for something that transforms points represented as vectors in three dimensions in such a way that distances between points are preserved.

Figure 5.9 can be used to visualise the situation. The concept “3D vector” might be found to begin with; then, looking at concepts that use “3D vector” in their definition, one might find “3D linear operator” and verify that this is indeed something that transforms three-dimensional vectors. Specializing iteratively, the researcher would eventually reach “3D orthogonal matrix” which is the appropriate concept.

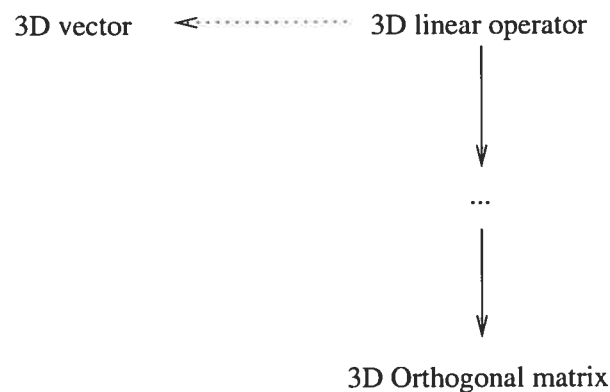


Figure 5.9: Some concepts and their interrelations in a navigable synthesis ontology. Full arrows denote specialisation; dotted ones denote definitional dependence.

Note that in both search situations presented, some amount of trial and error is to be expected, especially if the user is entering unfamiliar territory. Also note

that the user can be expected to assimilate new knowledge in the process of using of the system, which is valuable in itself and will make further uses of the tool more efficient.

5.5.2 Growing the ontology

I now turn to the important question of how to build the ontology. I propose a process that is inspired by the perspective put forth in the area of knowledge engineering by Tennison and Shadbolt [TS98], who describe ontologies as dynamic, living things that grow under favorable conditions as a diverse set of users comes into contact with it.

Tennison and Shadbolt introduced APECKS, a collaborative formal ontology construction environment for domain experts that is based on an internet-accessible multi-user text-based virtual environment named MOO (Multi-user domain - object-oriented [CN93]). APECKS' central feature is the presence of *roles* or personal ontologies expressed in a frame-based system [KMG95]. The roles that users construct reflect their own view of the domain. The construction of the ontology is carried out by users who expand the ontology and modify it. A comparison between the navigable synthesis ontology and the APECKS system is presented at the end of this section.

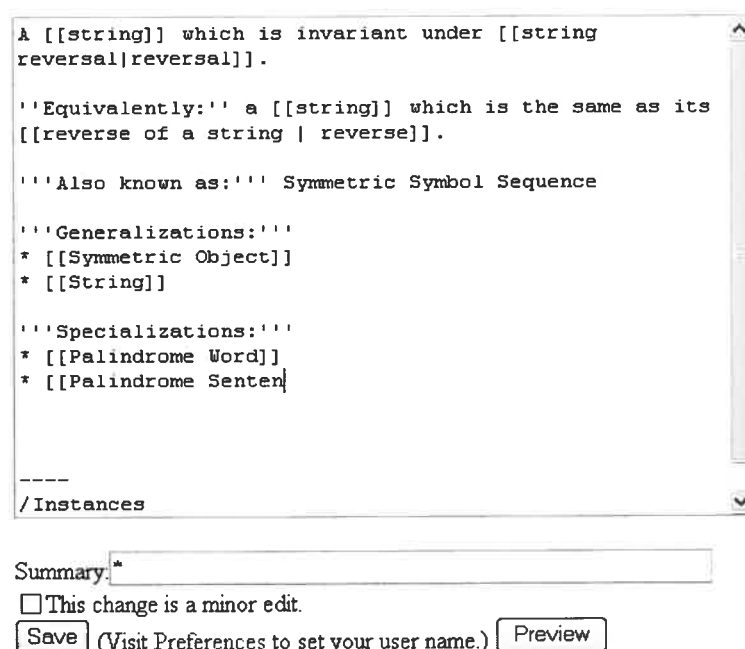
The ontology that I propose could never be considered complete; it is a perpetual work-in-progress. However, it can serve as a useful resource even as it grows. Were it not the case, the ontology could not attract many contributors.

The growth process results from individual contributions to the content and interindividual communication. Any collaborative creation system must accommodate both of these activities.

Content contribution. In my context, a content contribution takes the form of adding a concept or supplementing a concept description with additional information. Small investments are possible: even to add a single link to a new specialisation or a new generalisation of a concept (as illustrated by Figure 5.10) is a worthwhile contribution, which may be enough to stimulate activity around that concept. When multiple users are “present” at the same time, quick feedback is possible thanks to the recent changes feature, which makes the process more interactive and stimulating for participants.

Editing Palindrome

[Unifying Concepts Wiki \(home\)](#) | [Recent Changes](#) | [Preferences](#)



```

A [[string]] which is invariant under [[string
reversal|reversal]].

''Equivalently:'' a [[string]] which is the same as its
[[reverse of a string | reverse]].

''Also known as:'' Symmetric Symbol Sequence

''Generalizations:''
* [[Symmetric Object]]
* [[String]]

''Specializations:''
* [[Palindrome Word]]
* [[Palindrome Senten]

----
/Instances

```

Summary:

☐ This change is a minor edit.

(Visit [Preferences](#) to set your user name.)

Figure 5.10: Adding a specialisation called “Palindrome Sentence” to the “Palindrome” concept page.

It should be noted that many concepts do not yet have standard names — they either refer to implicit knowledge that usually goes unnamed, or they are genuinely new concepts. These names simply have to be invented. This allows an escape

from the limitations of the established vocabulary for any one domain.

An important consideration is how much rework is necessary if a concept in the ontology is found to be unsatisfying. The answer can be very different depending on whether definitions are revised in-place or if the new definitions appear under new concept names. In the first case, every concept that (directly or indirectly) relies on the revised concept in its definition is potentially affected by the change. The second strategy should be favored because it does not directly impact the remainder of the ontology; after the modification, it suffices to change the references that should point to the revised concept.

Interindividual communication. Interindividual communication can be broken down into two types: communication about the content of the pages proper, and communication about other matters (e.g., meta-discussion on methodology). Individual discussion pages that are subordinated to concept pages allow focused asynchronous discussion of the content, as was exemplified by Figure 5.5. To discuss, people simply take turns writing into the page, replying to one another.

A separate Wiki may also be set up as a space for meta-discussion and to let contributors set up personal pages. Such an approach is already being used and functions well in the context of the Wikipedia project.

Comparison with APECKS. There is commonality between the development model here and that presented by Tennison and Shadbolt. The core idea is similar, in that both approaches are designed to enable people who are not trained as knowledge engineers to collaborate on and communicate about the development of an ontology.

There are however significant differences. First, a navigable synthesis ontology is shared among users and doesn't explicitly feature roles, so that divergences be-

tween participants' views become more immediately apparent. Since the ontology is unlimited in space, it is however possible to implement the equivalent of roles if participants consent to using different names for their concepts.

Second, the ways in which ontology development is made accessible to non-knowledge engineers differ. A navigable synthesis ontology uses natural language to express concepts and its free form sets no restriction on how the ontology may be shaped. On the other hand, the APECKS system is based on a formal representation but adds a layer to make the information more palatable to users; this structured layer, which has to be designed by knowledge engineers, places restrictions on ontology development — for instance, users cannot define new axioms. Thus, the system depends on the involvement of knowledge engineers.

Aside from this greater flexibility, a significant advantage of a navigable synthesis ontology over a system like APECKS is simplicity. In APECKS, users have to contend with arguably burdensome structures from the beginning. Also, the reconciliation of the various participants' ontologies depends on a similarly complex process. A navigable synthesis ontology is less formal and constraining, and can thus be less offputting to potential contributors. This is an important factor to consider for a tool that aspires to a broad and diverse user base.

5.5.3 Using the ontology to locate knowledgeable people

Each concept in the ontology is not only a source of knowledge; it is also in effect a meeting point. People who work on the same concepts have an element of shared culture. The ontology, provided that it identifies who has worked on what, enables them to find one another. Since the concepts are very specific, there can be a close match in thinking between individuals who connect in that way. These links feed back into the knowledge networks of the people involved.

5.6 Limitations of synthesis ontologies

5.6.1 Limitations of my implementation

Here I discuss the limitations of my implementation of a navigable synthesis ontology, and possible ways to escape them.

First, as I have already indicated, my implementation is based on standard hypertext (following the HTML standard), which is the most common but perhaps not the ultimate way to navigate a conceptual web. Should a superior support for knowledge (virtual reality?) be devised, migration might be possible or not, depending on the availability of means to translate both the text and the linking mechanisms into the new medium.

Second, participation depends upon access to the Web. Although the web is currently nowhere near ubiquity, virtually all scholars in rich countries can access it, but many scholars outside these countries have limited access. The Internet population is however growing fast — estimates put the number of people online at 1 billion by 2005 [IR], which suggests that this hurdle to accessibility will soon be overcome.

Third, it is not clear how many people might want to use the system if it develops; this number might exceed the practical scale for maintaining the centralised architecture that I use. A decentralised architecture (such as exists on the Web itself) might be needed for a very large-scale project.

Fourth, as the ontology does not use a machine-readable language, there is potential for human error. Specifying concepts in a formal language would enable automatic verification of consistency; on the other hand, it would make it harder to work with (imagine writing research papers in completely formal language).

5.6.2 Limitations of the general design

Here I discuss the limitations that seem inherent to the general idea of a navigable synthesis ontology.

The first limitation has to do with the ways of thinking that using the tool invokes. Exploration as described in Section 5.5 requires two essential ingredients: that the user think as clearly as possible about what he is seeking, in such a way as to have the ability to choose wisely where to go from each page; and that he be prepared to encounter the unknown and learn as he goes in order to find his way around unfamiliar regions of the conceptual web.

While goal orientation and the ability to learn are defining characteristics of the human species, it is not clear whether every potential user can tap into these abilities to a sufficient degree to benefit from the tool. In essence, we're confronted with the core human issues of learning. A navigable synthesis ontology obviously does not make those issues vanish by its mere presence; if anything, its simplicity squarely brings them to the fore.

The second limitation has to do with the ways of thinking that building the ontology simultaneously requires. First, a mathematical or logical way of thinking is necessary, because every connection in the ontology must be rigorously stated and justified. In other words, the ontology requires that all knowledge be made explicit. On the other hand, a "right brain" or analogical way of thinking is needed in order to be able to generate new connections between things that are not yet formally linked. This makes the construction endeavour challenging. Then again, such a challenge is present to a certain degree in all design tasks that are founded on explicit logic.

5.7 Suitability for interdisciplinary communication

I posit that navigable synthesis ontologies provide a unique possibility for connecting knowledge across disciplines. The reason for this is that they provide an orderly means for people to represent and share whatever analogies they may find between concepts in different disciplines.

Consider two different concepts that are seen to be analogous. Then there are, on the one hand, some particularities that differentiate them, and on the other hand, characteristics that they have in common. Abstract out whatever it is that distinguishes them. The resulting concept is a generalisation of both of the original concepts. If the original concepts relate to different domains, this common generalisation may not have been identified in either domain. It may however occur in a third domain; or it may be completely novel. In any case, we now have a single abstraction that captures the commonality.

Two analogous concepts are then seen as having a common ancestor in some inheritance hierarchy. The closer the analogy, the less distant the common ancestor. The ancestor can itself be analogous to other concepts, entailing the existence of an even more abstract ancestor; and so on. Finding common generalisations is the basis for knowledge reuse, since everything that can be inferred with respect to the general concept is equally valid for any of its specialisations.

Now, as science popularisers and effective teachers know very well, it is considerably easier to get acquainted with a new idea when one is familiar with an analogous idea. Someone who knows about contagion will grasp the essentials of computer viruses more quickly than someone who does not, especially if reference is made to the former in explaining the latter. Knowing which analogies to use with a particular audience is a key to conveying one's ideas effectively.

In a navigable synthesis ontology, a concept of “generic virus” could be linked³ to both that of “computer virus” and that of “biological virus”, providing an easy-to-cross bridge between disciplines. Another link could be made between the software engineering notion of a “scenario” and the artificial intelligence idea of a “case” (in case-based reasoning), which have a common generalisation that might be called “problem”. These are but two examples. A well-developed ontology will feature a multitude of such bridges, each one providing a unique window for people on both sides of a formerly thick language barrier, which will appear to them when they seek it.

To sum up, my hypotheses here are that 1) analogies are powerful ways to make knowledge accessible across disciplines; and that 2) navigable synthesis ontologies enable analogies to be organised in such a manner that finding concepts analogous to a given concept becomes easier. Insofar as these hypotheses hold, such ontologies constitute a medium that is uniquely suitable for interdisciplinary knowledge sharing.

³Possibly through several degrees of specialisation

Chapter 6

Assessing the effectiveness of the tools

*Aristotle maintained that women have fewer teeth than men;
although he was twice married, it never occurred to him to verify this statement
by examining his wives' mouths.*

Bertrand Russell

Les grandes personnes aiment les chiffres.

Antoine de Saint-Exupéry

This chapter is concerned with the goal of assessing the effectiveness of the tools presented in the three previous chapters. I do so through case studies and surveys conducted with people who have used such tools.

Section 6.1 offers preliminary considerations on the methodology. Section 6.2 assesses the usefulness of personal knowledge sharing for communicating knowledge across disciplines. Sections 6.3 and 6.4 do like-

wise for open shared knowledge repositories and navigable synthesis ontologies.

I now turn to the question of evaluating whether the tools that I have introduced in the three previous chapters are helpful in enabling knowledge to move across disciplines. To this end I use case studies for all the tools under consideration, and user surveys for the first two. The case studies mention several different individuals and it is possible to get lost among the names. For this reason, Appendix A gives a brief “Who’s who”-type listing of people involved in the case histories.

6.1 Notes on methodology

As the reader probably expects, evaluating whether the tools that I have introduced are helpful in enabling knowledge to move across disciplines is not something that readily lends itself to measurement in an undisputably objective, quantitative manner. As Perkmann points out [Per02], when they exist, quantitative measures relating to knowledge systems are seldom meaningful and unambiguous.

Perkmann rather says we should “make a convincing case based on good arguments, pilot projects and case evidence and use quantitative tools as a supporting heuristic.” This is the approach that I use in this work. I have already given arguments of a theoretical nature for the usefulness of the methods I am advocating in Sections 3.5, 4.5, and 5.5. In this chapter I offer empirical evidence in the form of case studies for all three tools, and the analysis of surveys on the first two tools.

6.2 Personal knowledge publishing

My aim is to gather and look at evidence weighing for and against the hypotheses that I have put forth in Chapter 3. Those hypotheses were to the effect that:

1. Personal knowledge publishing, as a practice, allows researchers to establish *meaningful personal contacts* with researchers in fields other than their own.
2. Personal knowledge publishing allows *knowledge to be shared across fields*.

6.2.1 Case studies

In this section, I examine the experiences of three scholars who have attempted to use personal knowledge publishing for the purpose of sharing knowledge with people outside their discipline. The first is Lawrence Lessig, professor of Law at Stanford University. The second is Lilia Efimova, a Ph.D student in knowledge management who is at Telematica Institut in the Netherlands. I am the third researcher. Unsurprisingly, I was able to provide more detail on my own experience.

6.2.1.1 Lawrence Lessig

Professor Lawrence Lessig is a world-renowned expert on constitutional law as it applies to cyberspace. The best way to introduce him is to quote from the biography on his personal website¹:

¹URL: <http://cyberlaw.stanford.edu/lessig/bio/short/>

Lawrence Lessig is a Professor of Law at Stanford Law School and founder of the school's Center for Internet and Society. Prior to joining the Stanford faculty, he was the Berkman Professor of Law at Harvard Law School. Lessig was also a fellow at the Wissenschaftskolleg zu Berlin, and a Professor at the University of Chicago Law School. He clerked for Judge Richard Posner on the 7th Circuit Court of Appeals and Justice Antonin Scalia on the United States Supreme Court.

More recently, Professor Lessig represented web site operator Eric Eldred in the ground-breaking case *Eldred v. Ashcroft*, a challenge to the 1998 Sonny Bono Copyright Term Extension Act. Lessig was named one of Scientific American's Top 50 Visionaries, for arguing "against interpretations of copyright that could stifle innovation and discourse online."

He is the author of *The Future of Ideas and Code and Other Laws of Cyberspace*. He also chairs the Creative Commons project. [...]

Lessig teaches and writes in the areas of constitutional law, contracts, comparative constitutional law, and the law of cyberspace.

The Lessig News Log. The Center for Internet and Society hired Kathryn Yu as web guru in October 2001, and she was put in charge of his web presence. At that time she created a weblog, called the *Lessig Log*, that she has since been using to post news about Lessig's activities and related press. Since Lessig did not directly participate in writing this weblog, it does not qualify as an instance of personal knowledge publishing.

The Lessig Blog. Lessig started a personal weblog of his own on August 20, 2002. What initially prompted him to do so was that he wanted to engage in public debate with software developer (and weblog pioneer) Dave Winer.

Over the following months, he used his weblog mainly to discuss political and legal ways to counteract excessively restrictive regulation on technology and copyright.

Figure 6.1 is a sample post from the *Lessig Blog*.

The Silent 5: II

There's a growing and interesting thread at the *Volokh Conspiracy* about the Lopez argument that we made. Glenn Reynolds had a sensible post on the tension between strict constructionists (or as I have called it, the silent 5) and the result in *Eldred*. Juan non-Volokh agrees, disagreeing with Orin Kerr.

Orin argues that Lopez/Morrison were federalism cases; *Eldred* was clearly not. That's no doubt true, but missing from the opinion in *Eldred* is an explanation why enumerated powers get limited in the context of federalism, but not elsewhere. Judge Sentelle couldn't find such a reason. Maybe there is one. But the principle that would justify limiting power in one context but not in another should have been articulated.

But Juan says that the "strict textualist argument" that Glenn advanced was not advanced by us. He says this argument was Erik Jaffe's. It is true that Jaffe made this argument in the Court of Appeals. But it is not true that we failed to make a strict constructionist argument in the Supreme Court. Indeed, a section of our brief expressly argues that the "grant of power" was the "to promote progress" clause, and that that clause is not a "preamble."

The only difference between Jaffe's position and ours was that Jaffe's would have authorized a court to evaluate any copyright act to test whether it "promoted the progress of science." We thought that was too aggressive a position to take (on at least this point we were right!). Our argument instead was simply that the grant of power must at least be used to interpret the scope of "limited times." That while it was not an independent substantive constraint, it should be used to interpret the scope of the power. This is more "textualist" than Lopez itself: Lopez grounds its *reductio* on a background view about federalism; we grounded our *reductio* on a view about "to promote the progress of science."

Eugene passes on Eugene Kontorovich's view that this case was really about original-congress interpretations — that the silent 5 could vote as they did because they were simply ratifying what the original congress did. That is indeed the most charitable read, though again, it is just bad history. Abstracting the fact that in 1792 the framers had not yet fixed on what it would mean for a law to be unconstitutional, in 1790, there is zero evidence that the framers would have believed the *extension* of an existing term was within Congress's power. For the reasons argued extensively by the historians, Stevens, and us in our brief, the 1790 Act was not that. Thus it may well be that this case was all about one-step originalism. But on that basis, it was poorly reasoned.

posted on Jan 10 03 at 9:52 AM | eldred.cc | 11 comments

Figure 6.1: Lessig Blog, January 18, 2003

Lessig also used his weblog to interact with several other scholars who also edit weblogs, among which:

- Dr. Edward Felten, a professor of Computer Science and Director of the Security Internet Programming Laboratory at Princeton University. Felten edits the *Freedom to Tinker* weblog².
- Alex Golub, a graduate student studying anthropology at the University of Chicago. Golub edits the *Golublog*³.
- Eugene Volokh, who teaches "free speech law, copyright law, and the law of government and religion" at UCLA Law School. Volokh co-edits the *Volokh*

²<http://freedomtotinker.com/>

³URL: <http://lancelot.uchicago.edu/log>

Conspiracy weblog ⁴.

- Dr. David Weinberger, a writer, consultant and prominent thinker on the Internet and related issues. Weinberger edits *JOHO the Blog*⁵.
- Dave Winer, a software developer and writer and fellow at Harvard University's Berkman Center for Internet and Society. Winer edits the *Scripting News* weblog and is leading, as of 2003, an initiative to introduce weblogs as a means of fostering the exchange of ideas between the various schools of Harvard University⁶.

With the exception of Eugene Volokh, the people listed above do not specialise in Lessig's area. As Lessig's interests lie at the intersection of law, technology, collaboration, innovation, and even history, his weblog became a good platform to interact with people with different backgrounds who could contribute to advancing thinking on the themes he cared about. For instance, his weblog was the center stage for a discussion of the transparency of object code with Dave Winer and others in August and September 2002.

On several occasions, Lessig also used his weblog to request (and receive) help in preparing the Eldred case in which he attempted to convince the American Supreme Court that Congress should not be allowed to indefinitely extend copyright terms.

Quantitative data

Here are a few quantitative indicators of personal connections and knowledge exchange that have taken place because of Lessig's publishing the *Lessig Blog* (obtained on February 20, 3003):

⁴URL: <http://volokh.blogspot.com>

⁵URL: <http://www.hyperorg.com/blogger>

⁶URL: <http://blogs.law.harvard.edu>

- Weblog statistics service Technorati reports 343 inbound weblogs (weblogs that have at least one link to the Lessig Blog) and 499 inbound links⁷.
- More than 500 public comments on posts have been left by visitors.

6.2.1.2 Lilia Efimova

Mathemagenic. Lilia Efimova, who is beginning Ph.D. research in which she is trying to bridge eLearning and Knowledge Management, started a weblog called “Mathemagenic” (meaning: “giving birth to learning”) ⁸ on June 21, 2002. On the reason for starting her weblog she wrote:

“I was also looking for something to make my learning easier. I had so many notes on pieces of paper, in files, in yellow outline of print-outs, in books, in collections of links of references, in my head... I needed one point access for collecting those notes, relating them, reflecting, sharing and discussing.”⁹

Over the following months, Efimova wrote about articles that she had read, being careful to explain how the content connected with her own research ideas. She also quoted several other weblogs or other online sources and commented on them, in several instances generating indirect discussions by way of successive posts.

In September, Efimova went to the Knowledge Management Summer School at Sophia Antipolis and took numerous notes of the discussions in her weblog. She did likewise two months later, when the Quaerere group (formed of knowledge

⁷URL: <http://www.technorati.com/cosmos/links.html?url=http://cyberlaw.stanford.edu/lessig/blog/>

⁸URL: <http://blog.mathemagenic.com>

⁹*Mathemagenic*, June 21, 2002. URL: <http://blog.mathemagenic.com/2002/06/21.html>

management Ph.D. students) met in Brussels ¹⁰. Figure 6.2 is an example of these writings¹¹ from *Mathemagenic*, which exemplifies the content found in this personal publication.

■ **Q-Dialog: confused thoughts that bring new lights**

...we all have in common the need to broaden and to fundament our ideas about how we view Knowledge Management as a field. We can do this thought reading and attending conferences though there is a level, which neither books nor lectures may address, and that is the dynamic process of freely expressing less clear and even confused thoughts that bring new lights and new perspectives through their communication with the reasoning of others [Angela's process story]

We discussed yesterday night that best ideas seems to emerge from total mess and it seems that you need that mess to have ideas. What if we exploit it: we can have a session discussing not our findings, but our confusions or mistakes. This should help to move from presenting and refining existing ideas to supporting the emergence of new ideas.

categories: [Quaerere](#)
react: [comment \[0\]](#)

■ **Q-Dialog: content and process, reflections and suggestions for the future**

[Sorry if this is not clear. It's because the context is in my head. Probably, reading about [content and process stories](#) can help to understand.]

It's nice to have a laptop with me: 4,5 hours by train give a lot of time for thinking. I'm reading Angela's process story and thinking about it, schemas, and our passion to discuss process rather than content at Quaerere Dialog:

- It seems that Angela and Geri talk about very similar things using different languages. But Geri has something down to the earth: free discussion as a method to challenge our schemas.
- Somehow what they research is very close to what we want to do in Quaerere group: **double-loop learning**, reflecting on how we come to our ideas as well as challenging and changing our assumptions. If others agree with it, it would be easier to search for a method to organize our next meeting.

So, let's assume that we would like to focus on reflection upon our experiences. In the research context experience is an evolution of ideas. Ideas are content, process is how and why they evolve. So, there is one question that is teasing my mind: can we focus on process without content?

Figure 6.2: Mathemagenic, December 15, 2002.

Through her weblog, Efimova made contact with a number of researchers, among whom¹²:

- Thomas N. Burg, who does research on knowledge networks and online communities. Burg edits the *Randgänge* ("learning at the edge") weblog¹³

¹⁰URL: <http://blog.mathemagenic.com/2002/12/13.html>

¹¹URL: <http://blog.mathemagenic.com/2002/12/15.html>

¹²A systematic index of people referred to in this chapter can be found in Appendix A.

¹³URL: <http://randgaenge.net/>

- Sebastian Fiedler, an instructional design specialist and Ph.D. candidate in the area of educational technology. Fiedler edits the *Seblogging* weblog¹⁴.
- Jim Flowers, who is a professor of online education at Ball State University and edits *Jim Flowers' Radio weblog*¹⁵.
- Jim McGee, who researches information technology and knowledge management at Kellogg School of Management. McGee edits a weblog called *McGee's Musings*¹⁶.
- Sébastien Paquet, the author of this thesis, who researches scholarly communication and knowledge management and edits *Seb's Open Research*¹⁷.

Efimova also used her weblog to carry out conversations with Ton Zijlstra¹⁸, who is doing research on the philosophy of science at the university of Twente in the Netherlands. Efimova didn't know any of these people before she started publishing her knowledge on Mathemagenic.

It must be noted that links have value for people at both ends; so reciprocally, the people mentioned above benefited from gaining a connection to Efimova.

Many fruitful ideas have been exchanged by Efimova and her newfound colleagues through this medium, as can readily be seen by perusing the archives ¹⁹.

Quantitative data. Here are a few quantitative indicators of personal connections and knowledge exchange that have taken place because of Efimova's publishing *Mathemagenic* (obtained on January 20, 2003):

¹⁴URL: <http://seblogging.cognitivearchitects.com>

¹⁵URL: <http://radio.weblogs.com/0113212/>

¹⁶URL: <http://www.mcgeesmusings.net/>

¹⁷URL: <http://radio.weblogs.com/0110772/>

¹⁸URL: <http://interdependent.blogspot.com>

¹⁹accessible through the calendar in the upper right-hand corner of the page

- Technorati reports 37 inbound weblogs (weblogs that have at least one link to Mathemagenic) and 75 inbound links ²⁰.
- Her comment monitor reports 74 public comments on posts in her weblog²¹
- Efimova reports receiving more than eighty substantial emails in connection to her weblog presence by about a dozen different people, most of whom she didn't know beforehand.

6.2.1.3 Sébastien Paquet

I now turn to my own experience with personal knowledge publishing, which I will narrate with an eye towards highlighting the turning points during that experience.

Seb's Open Research: A chronicle. I started a personal knowledge publication called "Seb's Open Research" in July 2002 ²². The subtitle was initially "Pointers and thoughts on the evolution of scholarly communication, collected by Sébastien Paquet" but I changed it to the more general "Pointers and thoughts on the evolution of knowledge sharing and scholarly communication, collected by Sébastien Paquet" in September of the same year.

At that point I knew of only one researcher who shared my interest in personal forms of academic knowledge sharing: Jill Walker, whom I had found through a Google query for "research weblogs".

²⁰URL: <http://www.technorati.com/cosmos/links.html?rank=links&url=http%3A%2F%2Fblog.mathemagenic.com>

²¹URL: <http://www.myelin.co.nz/commentmonitor/?blogname=Lilia&usernum=109961&url=http%3A%2F%2Fradiocomments.userland.com%2Fcomments&stype=rsc>

²²The weblog can be accessed at <http://radio.weblogs.com/0110772>

Starting off

I started off my weblog using the *push* strategy explained in Section 3.5, chiefly by citing posts from other weblogs and writing the thoughts that had occurred to me as I read them. For purposes of illustration, Figure 6.3 shows my first post, dated July 19, 2002.

What if knowledge management actually is a technology problem?

KM as a technology issue. Current thinking holds that knowledge management's problems come from too much focus on technology when the key problems are about organizational processes and practices. I've said as much myself on many occasions. But this formulation risks perpetuating the myth that problems are either organizational or technological. We know the real world isn't that simple, of course. We shouldn't contribute to the confusion by oversimplifying our discussion.

...To me K-Logs represent the most interesting recent effort to address this need with a simple solution available right now. They offer a starting point that a knowledge worker can understand and build from.

[McGee's Musings]

I have to agree. Technology and organization go hand in hand. When it is adopted and subsequently (mis)used, technology tends to shape organizational practices. If it's introduced at the right time in the right way, you get a kind of symbiosis between the newly introduced technology and the innovative organizational practices that emerge from its use. But too often technology is designed with idealistic assumptions. The result is a gap that is too wide between current and wished-for practices. Most people are cautious. They won't try to jump long gaps, especially if they have yet to see anyone do it.

So proper design of technology is a problem because of this need take existing practices into account. Making the best possible tool is useless if users won't adopt it. The challenge is to make a tool that is simultaneously compatible with current and future practices, and bridges between the two. It's not easy.

Figure 6.3: Seb's Open Research, first post, July 19, 2002.

Showing up on other radars

The first few days were pretty quiet; nobody had yet seemed to have noticed my presence. By July 29, I had gotten pretty comfortable with the notion of K-Logs (knowledge management weblogs) and wrote the following post.

The Case for using K-logs in Research

K-Logs are a close fit to the academic culture. Here are ten reasons why.

1. Scholars value knowledge. They have a lot of it to manage and track.
2. A scholar's professional survival depends on name recognition. A K-log can help provide visibility and recognition.
3. Scholars are used to writing; most of them can write well.
4. Scholars are geographically disparate. They need to nurture relationships with people that they seldom meet in person.
5. Scholars need to interlink in a person-to-person fashion (see [interlinktural](#))
6. Scholars already rely heavily on interpersonal trust and direct communication to determine what new stuff is worth looking at. Such filtering is one of the central functions weblog communities excel at.
7. For many scholars, the best collaborations come about when they find someone who shares their values and goals (this is argued e.g. in section 3 of Phil Agre's excellent [Networking on the Network](#)). The personal output that is reflected in one's weblog makes it much easier to check for such a match than work that is published through other channels.
8. Scholars recognize the value of serendipity. Serendipity can come pretty quickly through weblogging; see [Manufactured Serendipity](#).
9. Every scholar must strive to be a knowledge hub in his niche, and an expert in related areas. A K-log is a good medium for this, as it is a way of letting knowledge flow through you while adding your personal spin.
10. Scholars pride themselves on being independent thinkers. K-logs epitomize independent thought.

The above points hold whatever the field of inquiry. Actually they are some of the reasons why researchers created the Internet and the Web in the first place. Weblogs, as an evolved, living form of web home pages, simply increase the incentive to get involved.

However a number of issues might cause resistance to the adoption of weblogs by academics:

1. It takes time.
2. "The technology is not well-established and tested at this point."
3. Many people don't like being among the first ones doing something.
4. Not all scholars are used to the Web and hypertext.
5. Shyness and fear of public mistakes. Many scholars won't write unless they have to. They may especially be reluctant to publicly expose ideas that they haven't tested.
6. Fear that someone else will pick up their ideas and work them out before they do.

This being said, a few researchers have dared to start weblogging. The best list I have found is compiled by Jill Walker: [Research Blogs](#). See also the interesting article [Blogging Thoughts: Personal Publication as a Research Tool](#) by Mortensen and Walker.

(I intend to continue building the case in [this wiki page](#) as more ideas and links come by)

[comment \[1\]](#) 10/10/02 12:01 #

Figure 6.4: Seb's Open Research, July 29, 2002.

I contributed the contents of the above post on the K-Logs mailing list²³ and traffic started coming to my site twelve hours later, following a mention in David Carter-Tod's *Serious Instructional Technology* publication. Within a week the post had been cited in a dozen other weblogs.²⁴ This is how I appeared on the radar of a few other weblog authors.

New connections

At that point I started using the *pull* strategy of Section 3.5. Tracing the new links pointing to my weblog that appeared in my referer log enabled me to discover different people who had found me. Among those was Sebastian Fiedler, an instructional design specialist and Ph.D. candidate with several interests overlapping mine, notably the use of personal knowledge publishing in education.

Traffic increased gradually on my site as I continued posting my thoughts. By regularly consulting my referer logs, I realised that many researchers whom I didn't know had found me. As in the case of Sebastian, several of these had interests in common with me:

- Andy Edmonds, who researches human factors, Web usability and information architecture at Clemson University. Edmonds edits the SurfMindMusings weblog²⁵.
- Lilia Efimova, who is doing research at Telematica Instituut in the Netherlands, trying to build bridges between learning and knowledge management. Efimova edits the Mathemagenic weblog discussed in the last section²⁶.

²³URL:<http://groups.yahoo.com/group/klogs/>

²⁴A list of citations can be found at the following URL:
http://www2.iro.umontreal.ca/~paquetse/cgi-bin/om.cgi?The_Case_For_K-Logs_In_Academia/Backlinks.

²⁵URL: <http://www.surfmind.com>

²⁶URL: <http://blog.mathemagenic.com>

- Alexander Halavais, who is doing research in communication at Buffalo University. Halavais is interested in new forms of group communication, collective intelligence, and group-forming; he edits the Halavais News weblog²⁷.
- Spike Hall, from the School of Education at Drake University, who researches individual, group, and organisational phenomena. Hall edits the Connectivity weblog, and has started analysing personal knowledge publishing in terms of its learning potential²⁸.
- Elisabeth Lane Lawley, who researches information science at Rochester Institute of Technology. Among other things, she is interested in the possible social and educational impact of weblogs. Lawley edits the Mamamusings weblog²⁹.
- Tom Munnecke, who does research as a Visiting Scholar at the Stanford University Digital Visions Program. Munnecke asks the question, “What is the simplest thing I can do which will have the maximum global humanitarian uplift?” Very similarly to me, he thinks in terms of cooperation networks, self-organisation, dealing with information overload, and online communities³⁰.
- Laura Trippi, a researcher at Simon Fraser University with an interest “in networked processes and the ways in which they generate, circulate, and preserve meaning, in understanding culture as a complex system”³¹.

Pursuing my *exploration* efforts as per Section 3, I found in these newly discovered weblogs yet more references to other research weblogs whose authors I found close affinity with:

²⁷URL: <http://alex.halavais.net/news>

²⁸URL: <http://radio.weblogs.com/0106698>

²⁹URL: <http://www.it.rit.edu/~ell/mamamusings>

³⁰URL: <http://www.munnecke.com/blog>

³¹URL: <http://www3.bc.sympatico.ca/latrippi/netvironments>

- that³² of Mark Bernstein, the chief scientist at Eastgate Systems, Inc., who develops the personal content management assistant and weblog tool Tinderbox. Bernstein is also the person behind the Weblog Kitchen knowledge repository, which I described in Section 6.3.
- that of Stephen Downes, an e-learning researcher at the National Research Council of Canada in Moncton, New Brunswick, Canada. Downes issues the OLDaily (Online Learning Daily)³³ every weekday on his web site. One of our common interests is open approaches to knowledge sharing and learner-directed learning.

It is very probable that I would never have found out about most of these people and assessed our commonality of interest were it not for our respective involvement in weblogging. I had never come across publications by any of the above prior to finding their weblog. Even if I had, it wouldn't have told me what they are currently interested in. Moreover, as was observed in the previous case, the connections are bidirectional, so that these people have also benefited from their personal knowledge publishing activity by connecting to me.

Conversations

Through my weblog, I gradually found myself reading most of the weblogs mentioned above on a regular basis. I participated in many indirect conversations (about two or three a week) involving the people mentioned above as well as some others on a variety of topics related to my research. For instance, one of these conversations, involving Pete Harbeson, Alison Fish, Ron Lusk, and Al Macintyre, was recapitulated by Phil Wolff in "Why Johnny Can't Klog"³⁴. Another one involved Seth Russell, Sebastian Fiedler and me³⁵.

³²URL: <http://markbernstein.com>

³³URL: <http://www.downes.ca>

³⁴URL: <http://dijest.com/aka/2002/09/02.html#a1978>

³⁵URL: [http://seblogging.cognitivearchitects.com/discuss/msgReader\\$410](http://seblogging.cognitivearchitects.com/discuss/msgReader$410)

For purposes of illustration, Figure 6.5 reproduces yet another one, this time involving Roland Tanglao, Phil Wolff, John Robb, Lilia Efimova and me, which helped better work out my ideas about the place of knowledge repositories, as it was captured by Efimova ³⁶:

 Tuesday, August 20, 2002

• Taxonomies to digest knowledge in klogs

Via [Seb's Open Research](#): discussion about [Klogging roles](#) turns into discussion about [The natural progression for knowledge: from K-logs to well structured forms](#).

Blogs is definitely worse than simple discussion board to give proper credits to all who participated. I would just cite some ideas:

[Roland Tanglao](#): K-Log => (FAQ or other knowledgebase article) => directory.

[a Klog apart](#): Self-review is a powerful tool for learning. Going over my own posts for the past week, month, and quarter has shown patterns I missed, ideas I was skirting but never wrote outright. It reinforced brief social connections, blogs to which I linked to and people with whom I briefly corresponded. It takes concentrated time and effort. It helps me to print out all the pages on my blog for that period; something about shuffling through paper.

[Seb's Open Research](#): Phil has a point here. But I'm afraid that even expert authors are seldom able to fight entropy in the manner described, unless they have plenty of time and motivation to do synthetic work. In the academic world the ratio of research paper authors to survey/textbook authors is perhaps 50:1. But in time, as the overall quantity of knowledge grows and grows, ultimately almost everyone will feel lost and the usefulness of "mappers" ought to be better recognized.

I treat blog as my external brain more than a publishing tool (at least so far). Could you imagine someone trying to digest knowledge stored in several brains to get something meaningful out of it? Could be funny :)))

From another side, there is at least one editor who might want to digest posts in blog - it's me. At first, blog helps me to capture ideas by [articulating them](#), but then I want to look for patterns and connections that emerge. This would be something to do for the smart tools.

Date, time and categories are definitely not enough for me. I want something like this:

However, it is possible, once a K-Logging culture is in place to utilize taxonomy tools (tools like Wikis and Traction Software) to organize K-Log generated information into a larger whole. The key to success is to first lay the groundwork with a K-Log network and then leverage it after it begins to produce results. K-Logging puts the knowledge into a format that makes it easier to manipulate by a taxonomy tool. Longer term, I think most organizations will use combinations of the two types of tools to turn the Intranet into a rich, vibrant, and growing knowledge repository. ([John Robb on K-Logs and Taxonomies in Yahoo! Groups: klogs](#))

categories: [all KM](#) [k-logs](#)
 feed: [comment \[0\]](#)

Figure 6.5: Example of a multi-party conversation on knowledge logging (Mathemagenic, August 20, 2002)

Weblog-aided dissemination

On October 3, 2002, I put a 15-page research article titled "Personal knowledge publishing and its uses in research" online [Pq02] and wrote the post reproduced

³⁶URL: <http://radio.weblogs.com/0109961/2002/08/20.html>

in Figure 6.6 in my weblog to introduce it.

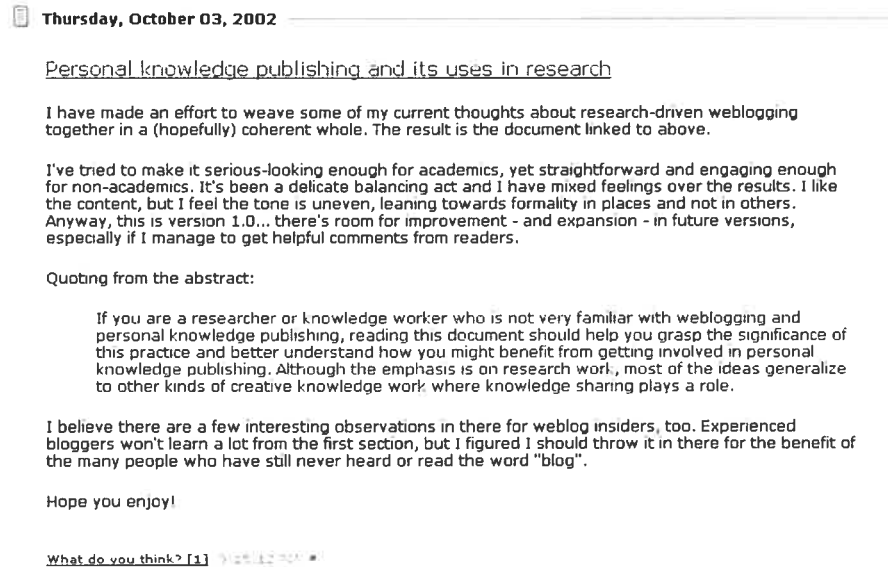


Figure 6.6: Introduction to the article “Personal knowledge publishing and its uses in research”, Seb’s Open Research, October 3, 2002.

Over the following months, the article was mentioned by more than 70 different people, of which about half wrote comments of some sort on the content (largely favorable, I am happy to report)³⁷. Subsequently, I was offered to republish the article³⁸ on the Knowledge Board (the portal of the European KM community) on December 16, 2002, and Dolores Tam kindly offered to translate the piece into French³⁹ so as to better spread the ideas to the francophone community. The text was also used for a course called “Writing for Digital Media” at the University of North Carolina⁴⁰.

³⁷A partial list of references to the paper, along with the date at which they were published, can be found at the URL :

<http://www2.iro.umontreal.ca/~paquetse/cgi-bin/om.cgi?>

[Personal_Knowledge_Publishing_And_Its_Uses_In_Research/Backlinks](http://www2.iro.umontreal.ca/~paquetse/cgi-bin/om.cgi?ap=1&id=96934)

³⁸URL: <http://www.knowledgeboard.com/cgi-bin/item.cgi?ap=1&id=96934>

³⁹URL: <http://radio.weblogs.com/0105068/stories/2002/12/23/traductionDePersonalKnowledgePublishingAndItsUsesInResearch.html>

⁴⁰URL: <http://www.unc.edu/~briman/jomc221.html>

Group-forming, squared

On October 9, 2002 I had an idea for a simple way of enabling weblog authors to cluster around shared interests and posted it to my weblog. The post was commented upon by John Robb (the president of UserLand, who develop the Radio web publishing software), Phillip Pearson, and Eric Hanson. Hanson thought the topic of group-forming was interesting enough that we ought to start a mailing list to discuss it.

We did set up a mailing-list based group on group-forming, and 60 people had subscribed within two weeks. The list attracted a wide diversity of participants, ranging from philosophers to sociologists, programmers, journalists, consultants and entrepreneurs. One month later we moved the list onto a collaborative weblog⁴¹ which was less active in terms of volume, but also more focused, as if the group had turned into more of a research team.

This group remains active to this day. Together, participants have discussed the nature of existing group-forming systems, theorised on the notion of finite and infinite groupspace, identified a subset of groups we call “purposeful groups” or teams, and discussed the problem of eliciting and matching individual motivations, among other things.

Phillip Pearson implemented my original group-forming system idea in the form of the Internet Topic Exchange⁴², which went live in mid-January 2003. This system has been in continuous use since that time and is growing in popularity among weblog authors⁴³. Phillip Pearson (with whom I have never actually spoken) and I have coauthored a paper on that experiment that we have submitted for a conference in February 2003 [PP03].

⁴¹URL: <http://aquameta.com/gf>

⁴²URL: <http://topicexchange.com>

⁴³As of February 28, 2003, 69 groups have come into existence and more than 600 posts have been submitted into the system

Blog research

Communication researcher Alex Halavais and information technology / library science researcher Elizabeth Lane Lawley showed up on my radar at about the same time, in October of 2002. They had a common interest in the social implications of weblogs, but they didn't know one another. I played the role of matchmaker by commenting on a post by Lawley on my own weblog, which was read by Halavais and led to him visiting Lawley's weblog and expressing interest in exploring common research interests.

Following that, Elizabeth Lane Lawley created a collective weblog called "blogresearch" to bring together people with a research interest into weblogs, and invited Halavais and a few others to become editors. A grant application to the U.S. National Science Foundation came out of this a few months later. This is another illustration of how weblogs can enable new connections between researchers in different fields.

Quantitative data

Here are a few quantitative indicators of personal connections and knowledge exchange that have taken place because of my publishing *Seb's Open Research*:

- Technorati reports 166 inbound weblogs (weblogs that have at least one link to mine) and 345 inbound links⁴⁴.
- my comment monitor reports 167 comments on posts in my weblog⁴⁵.
- I have exchanged over three hundred substantial emails in connection to my weblog presence with more than 30 different people in the last five months.

⁴⁴URL: <http://www.technorati.com/cosmos/links.html?rank=links&url=http%3A%2F%2Fradio.weblogs.com%2F0110772>

⁴⁵URL: <http://www.myelin.co.nz/commentmonitor/?blogname=Seb&usernum=0110772&url=http%3A%2F%2Fradiocomments.userland.com%2Fcomments&stype=rsc>

(For comparison, prior to that I had almost no connections with people outside my institution and received about one substantial email per month.)

6.2.1.4 Analysis

The three examples studied above indicate that it is possible to make personal contacts and share knowledge across disciplines using personal knowledge publishing.

I believe that the foregoing exposition illustrates rather clearly that personal knowledge publishing was instrumental in helping me establish new meaningful connections and exchange knowledge with several people outside by area of specialisation (which may be described as scholarly knowledge management/collaboration). After five months, I can confidently state that at least a couple dozen new people know me well. They are aware of what I'm interested in and of the quality of my work enough to send like-minded people my way if they encounter them.

I obviously do not intend to discontinue my personal knowledge publishing experiment. My weblog has helped me establish a public reputation as a serious thinker in a few months' time, even though I started out as a complete unknown. My publishing activity has even resulted in two organisations (the National Research Council (NRC) of Canada and Tomoye, Inc.) approaching me about possible employment. Moreover, in February, 2003, I was invited to the "Canada 2025 workshop" workshop, a two-day brainstorming session organized by the NRC, bringing together a few dozen future-oriented Canadian thinkers. Interestingly, prior to getting this invitation I had never had any contact with the organizers. They learned about me via my weblog and decided to invite me based on its contents.

My weblog has become a central part of my professional identity, and keeps bringing me new interesting readers and collaborators, with diminishing effort as more and more other weblogs point my way. I am not currently aware of a better way to make contact with such people.

It should be noted that, in both situations studied, the researchers have an interdisciplinary orientation. As was remarked in the introduction, this makes it next to impossible to find capable people with a similar interest using traditional means of networking. So the non-negligible time investment in personal knowledge publishing is perhaps more warranted for interdisciplinary-minded researchers than for those who stay more within disciplinary boundaries and thus have easier access to the knowledge and people they need.

6.2.2 User survey

Over the months of December 2002 and January 2003, I conducted a survey which enabled me to obtain quantitative data on the usefulness of personal knowledge publishing and open shared knowledge repositories for knowledge work, and in particular for interdisciplinary communication.

I overtly asked for assistance in conducting the survey on *Seb's Open Research*, and obtained feedback from several people. Among those was VeerChand Bothra, a software developer from India who operates the BlogStreet weblog search service⁴⁶. Bothra offered to help and actually performed most of the HTML and database work required to set up the poll, in addition to hosting it and publicizing it on BlogStreet. This is another illustration of how personal knowledge publishing can enable a significant degree of collaboration between people who don't know one another beforehand.

⁴⁶URL: <http://blogstreet.com>

The survey was publicised through weblogs, wikis and mailing lists, as well as on the BlogStreet forum⁴⁷, and it was conducted online. Respondents visited a web page⁴⁸ listing 25 questions. Of these, 23 were multiple-choice questions and the remaining two were open-ended. Participants filled in the questionnaire and had to click the “submit” button when done.

6.2.2.1 Questionnaire

The questions that were asked follow. (Note that many questions were included to collect data for other studies on knowledge management and are not directly used in this work.)

1. Have you ever read a weblog?
2. How often do you visit weblogs?
3. How often do you visit weblogs to locate interesting work-related material?
4. What percentage of the time that you spend reading for professional purposes is devoted to reading weblogs?
5. Are there weblogs that you regard as being among the best sources of information on a particular topic?
6. In your opinion, are there interesting ideas or information to be found in weblogs that are hard to find in other media?
7. Did reading weblogs help you find people whose opinion you would be willing to trust on certain matters of professional interest?
8. Do you edit a personal weblog?
9. How often do you update your personal weblog?
10. Do you use your personal weblog to jot down personal ideas that relate to your professional activities?
11. Do you use your personal weblog to keep track of interesting material that relates to your professional activities?

⁴⁷URL: <http://blogstreet.com/forum>

⁴⁸URL: <http://blogstreet.com/survey/weblog.html>

12. How often do you revisit past content (more than a week old) in your weblog for reference purposes?
13. Have you found people with professional interests close to yours by reading weblogs?
14. Have you found people with professional interests close to yours by looking at who linked to you, for instance using referer logs or weblog crawler data?
15. If you answered “Yes” to one of the previous two questions, do you think you could have connected with these people in another way than through weblogs?
16. Has your blogging activity helped you find people with professional interests close to yours but with a different professional background than yours?
17. If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through weblogs?
18. How often does weblogging help you fruitfully exchange ideas with people who specialize in the same area as you?
19. How often does weblogging help you fruitfully exchange ideas with people who specialize in a different area than your own?
20. How often do you use your weblog to ask questions?
21. If you’ve posted questions to your weblog, how many have resulted in useful answers from readers?
22. Professionally, you would describe yourself as? (a technologist, a journalist, a librarian, an educator, a lawyer, a consultant, a researcher, a knowledge worker but none of the above, not a knowledge worker)
23. How many years of professional experience do you have?
24. (Optional) In your own words, would you say that weblogging has changed the way you share and obtain work-related knowledge? In what way?
25. (Optional) What is your blog’s URL?

6.2.2.2 Results and analysis

The complete results of the survey are available on the World Wide Web⁴⁹.

⁴⁹URL: <http://www.iro.umontreal.ca/~paquetse/KnowledgeSharingSurvey.html>

In all, 177 people answered the survey. All but six had visited a weblog at some point. A hundred and forty people indicated their professional orientation. Of these, 23% described themselves as “a technologist”, 21% as “another kind of knowledge worker” (i.e. not one of the choices given), 18% as “a consultant”, 11% as “a researcher” and another 11% as “an educator”; the remaining 16% were distributed relatively evenly among journalists, librarians, lawyers, and people who were not knowledge workers. As regards professional experience, 29% had five years or less, 21% had from 6 to 10 years of experience, 33% had from 11 to 20 years, and 17% had more than 20 years of experience. Although I cannot tell for sure because of the way the question was set up, there were probably a few students among the 29% with five years experience or less.

In this section, I present and interpret the data that is relevant to evaluating the hypotheses that were put forth at the end of Section 3.5. I also investigate the relationship between the time investment and knowledge-related benefits of personal knowledge publishing.

Hypothesis 1: “Personal knowledge publishing, as a practice, *allows researchers to establish meaningful personal contacts* with researchers in fields other than their own.”

The responses to questions #7, #16, and #17 are relevant here. Figure 6.7 and Table 6.1 represent the distribution of answers to question #7, in the overall sample and in the subset of researchers (15 respondents). Note that the number of researchers is not large because personal publishing is only beginning to be recognised as a useful tool among researchers.

It is apparent from this data that trusting relationships can be established through the medium of weblogs, with 90% of all respondents and 87% of researchers answering positively to the question. 60% of researchers said they had found three

7. Number of people trusted on certain matters
of professional interest and found via weblogs

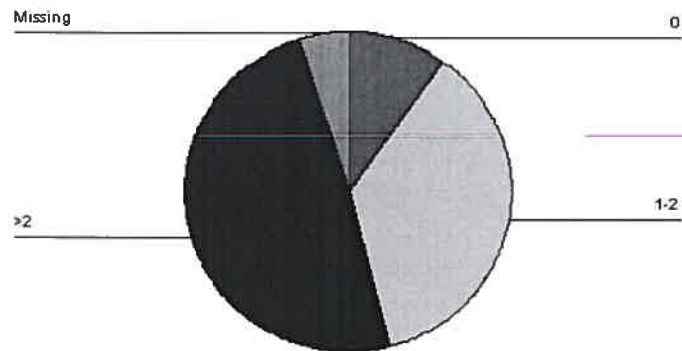


Figure 6.7: Distribution of answers to question #7, “Did reading weblogs help you find people whose opinion you would be willing to trust on certain matters of professional interest?”

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, >2	87	49.2	51.8	51.8
	Yes, 1-2	64	36.2	38.1	89.9
	No	17	9.6	10.1	100.0
	Total	168	94.9	100.0	
Missing		9	5.1		
Total		177	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, >2	9	60.0	60.0	60.0
	Yes, 1-2	4	26.7	26.7	86.7
	No	2	13.3	13.3	100.0
	Total	15	100.0	100.0	
Missing		0			
Total		15			

Table 6.1: Distribution of answers to question #7, “Did reading weblogs help you find people whose opinion you would be willing to trust on certain matters of professional interest?”

or more people they would be willing to trust on certain matters of professional interest via weblogs.

The next question that I consider examines the issue of interpersonal links across communities. Figure 6.8 and Table 6.2 show the distribution of answers to question #16, in the overall sample and in the subset of researchers.

**16. Has blogging helped you find people with
shared interests but with a different background?**

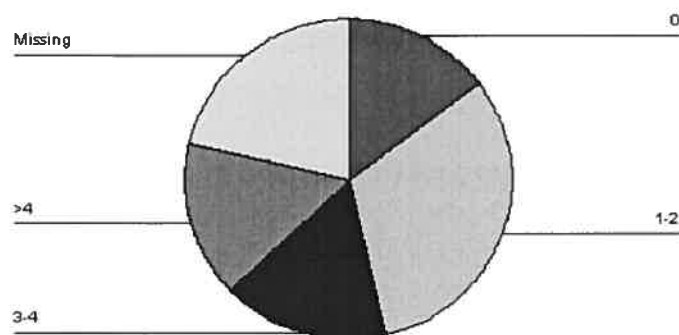


Figure 6.8: Distribution of answers to question #16, “Has your blogging activity helped you find people with professional interests close to yours but with a different professional background than yours?”

In both cases, 80% of respondents said they had been able to find people with a common interest but a different background, and 40% had found three or more such people, which indicates that weblogs do enable interpersonal links across communities of knowledge. (The average numbers of links reported were 2.54 and 2.37, respectively, for the general sample and the researchers sample.)

It is however possible that these links could have been established in other ways. Question #17 was designed to address this issue. Figure 6.9 and Table 6.3 illustrate the distribution of answers to this question, in the overall sample and in the subset of researchers.

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, >4	28	15.8	20.1	20.1
	Yes, 3-4	29	16.4	20.9	41.0
	Yes, 1-2	56	31.6	40.3	81.3
	No	26	14.7	18.7	100.0
	Total	139	78.5	100.0	
Missing		38	21.5		
Total		177	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, >4	2	13.3	13.3	13.3
	Yes, 3-4	4	26.7	26.7	40.0
	Yes, 1-2	6	40.0	40.0	80.0
	No	3	20.0	20.0	100.0
	Total	15	100.0	100.0	
Missing		0			
Total		15			

Table 6.2: Distribution of answers to question #16, “Has your blogging activity helped you find people with professional interests close to yours but with a different professional background than yours?”

17. Do you think you could have connected with
these people outside of weblogs?

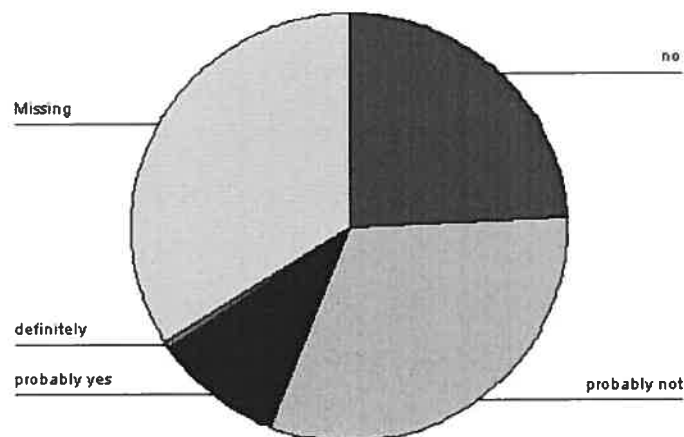


Figure 6.9: Distribution of answers to question #17, “ If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through weblogs?”

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	No	39	22.0	34.5	34.5
	Probably not	56	31.6	49.6	84.1
	Probably yes	17	9.6	15.0	99.1
	Definitely	1	.6	.9	100.0
	Total	113	63.8	100.0	
Missing		64	36.2		
Total		177	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	No	3	20.0	25.0	25.0
	Probably not	7	46.7	58.3	83.3
	Probably yes	2	13.3	16.7	100.0
	Definitely	0	0	100.0	
	Total	12	80.0		
Missing		3	20.0		
Total		15	100.0		

Table 6.3: Distribution of answers to question #17, “ If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through weblogs?”

Again both groups' answer patterns are similar. The respondents said in about 84% of the cases that they would not, or probably not, have made such connections were it not for their involvement in a personal knowledge publishing network. Personal knowledge publishing thus appears to be superior to other means of communication in this respect: it enables interpersonal connections that are difficult or impossible to establish by other means.

To summarise, the data collected in connection with these three questions.

Hypothesis 2: "Personal knowledge publishing *allows knowledge to be shared across fields*".

The responses to question #19 are relevant here. Figure 6.10 and Table 6.4 show the distribution of answers to this question.

19. How often does weblogging help you fruitfully exchange ideas with people outside your area?

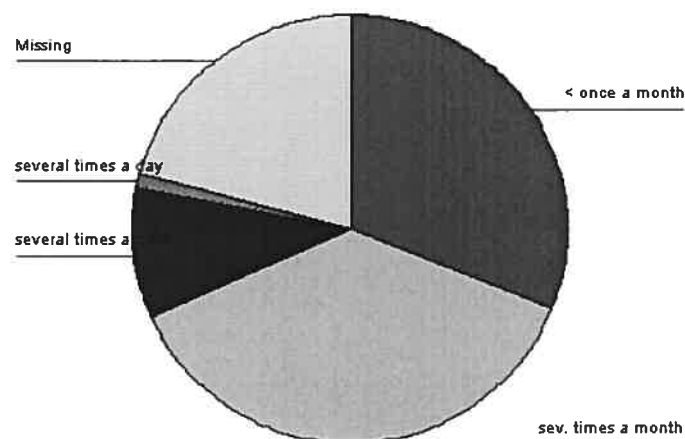


Figure 6.10: Distribution of answers to question #19, "How often does weblogging help you fruitfully exchange ideas with people who specialize in a different area than your own?"

In the overall sample, 60 percent of respondents claim that knowledge exchanges

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	less than once a month	55	31.1	39.3	39.3
	several times a month	66	37.3	47.1	86.4
	several times a week	17	9.6	12.1	98.6
	several times a day	2	1.1	1.4	100.0
	Total	140	79.1	100.0	
Missing		37	20.9		
Total		177	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	less than once a month	5	33.3	33.3	33.3
	several times a month	7	46.7	46.7	80.0
	several times a week	3	20.0	20.0	100.0
	several times a day	0	0	0	
	Total	15	100.0	100.0	
Missing		0			
Total		15			

Table 6.4: Distribution of answers to question #19, "How often does weblogging help you fruitfully exchange ideas with people who specialize in a different area than your own?"

with people outside their area occur several times a month or more. In the case of researchers, the corresponding figure is 66 percent. Judging from this distribution it seems reasonable to infer that personal knowledge publishing does enable ideas to circulate across communities of knowledge. Moreover, it appears that for those researchers, the weblog has displaced other media and become the channel of choice for sharing (some of) their ideas. This lends support to the second hypothesis.

Further support comes by way of some answers to the open-ended question #24, "In your own words, would you say that weblogging has changed the way you share and obtain work-related knowledge? In what way?". Five responses directly alluded to the better circulation of knowledge across disciplinary boundaries over the personal knowledge publishing medium; they are reproduced here.

One respondent indicated that weblogs help learn about people with different backgrounds:

More than work-related, blog reading has helped me learn quite a bit about people of different backgrounds, and cultures. In my opinion, it is the best way to get an up-close and personal view of the types of individuals you would never meet in real life. This alone makes the weblog an incredibly valuable resource.

Another one explicitly referred to creating interdisciplinary connections that he had previously been unable to make:

Yes it has; I now feel as if I have found other 'voices' with similar concerns... concerns that bridge disciplinary and geographic and even cultural boundaries.

A third respondent referred to diversity:

Absolutely... I've found it to be an extremely valuable way to communicate and collaborate with a diverse group of interesting and intelligent people.

A fourth respondent wrote (in French, translation follows):

“J’ai pu constater que le weblogging favorisait le “brassage” entre des communautés d’intérêts différents. Exemple : les diaristes “classiques” ont commencé à utiliser les outils de weblogging et cherché des conseils pour mieux les utiliser. Les blogueurs “techies” ont découvert en retour l’existence des journaux intimes.”

Translation: “I noticed that weblogging fosters a “mixing” between communities with different interests. Example: “Classical” diarists have started using weblogging tools and looked for advice on using them. Reciprocally, “techie” bloggers discovered personal diaries.”

A fifth response reads:

[...] “Unfortunately, I was not able to find any weblog dealing with the same or even similar topics (financial asset management) as mine. So, I shared a lot of work-related knowledge but obtained none. On the other hand, having the weblog proved to be a very interesting experiment for me and gave me the opportunity to connect with people in other professional fields (mainly internet/consulting/marketing etc.).”

Time investment and interdisciplinary connections. Recall that we asked the question of how much time one needs to invest in personal knowledge publishing to be able to establish connections with people outside one's field. I investigate this issue by examining the relationship between responses to question #4, which asked about the percentage of the time spent reading for professional purposes that was devoted to reading weblogs, and question #16, which asked for the number of people with shared interests but with a different background that were found through weblogs.

Figure 6.11 illustrates this relationship.

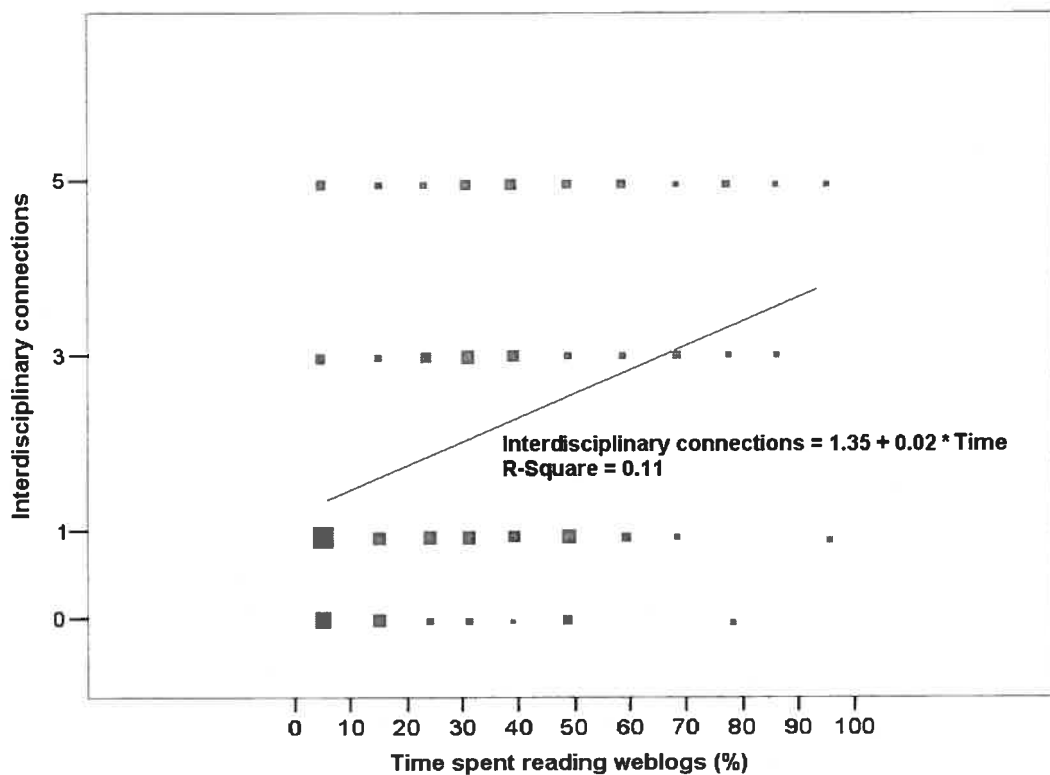


Figure 6.11: Percentage of the time spent reading for professional purposes that is devoted to reading weblogs vs. number of people with shared interests but with a different background that were found through weblogs. (Square area is proportional to response count.)

Respondents' ability to establish connections across interdisciplinary boundaries

Group	<i>N</i>	Mean	SD	SEM
General sample	124	1.44	1.02	0.09
Researchers	15	1.33	0.98	0.25
<i>t</i> value	0.367			
<i>p</i> value	0.714			

Table 6.5: Comparison of researchers and non-researchers on the number of interdisciplinary connections made through weblogs.

does indeed seem to increase slightly with use of weblogs. However, the linear regression does not indicate a very significant relationship ($R^2 = .11$), and the slope of the fit is not abrupt. It can be seen from the chart that several people were able to establish many connections in spite of investing comparatively little time in weblogging.

Difference between researchers and non-researchers. It can be interesting to look for differences between the subsample of researchers and the rest of the sample, which would indicate that the findings above obtained with the general sample might not apply to researchers. Table 6.5 compares the distributions of the “number of interdisciplinary connections” variable. This test does not show significant differences between the samples⁵⁰.

6.2.3 Evaluation criteria

Here I review the adequacy criteria that I presented in the introduction (Section 1.1.6), and examine the degree to which personal knowledge publishing fulfills them.

Accessibility. Weblog-type software is freely and commercially available on the

⁵⁰Note that the researcher sample is small. While differences don’t appear at this relatively coarse level of detail, it is not out of question that a larger sample would suggest a statistically significant difference.

Web. Installation and operation does not require much resources in terms of time or money, as the large number of weblogs by a diverse population illustrates⁵¹. In the context of research, access to personal knowledge publishing is limited to those researchers who can access the Web, but this is a large (and growing) proportion of researchers.

Appearance of value. The growing number of instances of personal knowledge publications by researchers indicates the value that they see in the activity. Though they clearly constitute a very small proportion of the total number of researchers, it must be kept in mind that none of them was pushed in any way to adopt the medium - each decision to use the tool was taken independently. Thus it can be said that by several independent opinions, the tool appears valuable enough to be tried.

Adequacy. The results of the survey presented above suggest that the tool adequately fulfils some of its users' knowledge sharing needs.

Inclusiveness. Anyone with access to the Internet can start a weblog and write whatever they like in it. There is no restriction to the ideas that can be published, other than what is legally forbidden, e.g. copyrighted or defamatory material.

Scalability. The weblog tool, by its decentralised nature, evidently scales as well as the World Wide Web itself.

Evolvability. Weblog technology is not rigidly standardised. Everyone is allowed (and welcome) to propose and implement extensions to the technology. Many people do, and as a result the weblogging environment has evolved significantly in the last few years, for instance with the deployment of RSS syndication (Section 3.4) and ecosystem mapping tools.

⁵¹Weblogs that are maintained by homeless people, such as Kevin Barbieux (URL: <http://thehomelessguy.blogspot.com/>), drive this point home in a more dramatic manner.

This concludes my assessment of the applicability of personal knowledge publishing to the interdisciplinary knowledge sharing problem. The next section investigates open shared knowledge repositories.

6.3 Open shared knowledge repositories

Here I wish to gather and look at evidence weighing for and against the hypotheses that I have put forth in Chapter 4. Those hypotheses were:

1. Open shared knowledge repositories allow researchers to build a community of people with an interest in an interdisciplinary, possibly ill-defined topic.
2. Open shared knowledge repositories allow researchers to build a useful knowledge base around an interdisciplinary, possibly ill-defined topic.

To this end, I will draw on two case studies of actual open shared knowledge repositories, and on a survey conducted with users of such repositories.

6.3.1 Case studies

6.3.1.1 Information architecture wiki

Information architecture is a nascent field of study that is concerned with “designing the structure of information spaces for human beings to work, play or learn in.” It has roots in several different fields including visual design, information design, library science, and engineering psychology (more commonly known as human factors).

The Information Architecture wiki (<http://IAwiki.net>) was created by Eric Scheid, an Australian information architect and database specialist, towards the end of 2001. As of the end of 2002, the wiki has grown to about 600 hypertext pages.

Several dozen people have contributed to the IA wiki, among which there are a number of people who do research in the field. A few of them are presented below:

- Mark Bernstein has a Ph.D. in chemistry but has been researching hypertext systems for a decade;
- Kat Hagedorn, Peter Morville, Tanya Rabourn, and Louis Rosenfeld are library science specialists who do research on information architecture;
- Victor Lombardi has degrees in journalism and music technology;
- Rashmi Sinha has a Ph.D. in cognitive psychology;
- Christina Wodtke studied art before becoming an information architect.

The variety of backgrounds in the above list reflects the interdisciplinarity of the field. Topics under discussion (which I cannot describe in detail here) include “Activity-centered design”, “Associative Navigation”, “Faceted Classification”, “Future Directions of IA”, “Hypertext Disorientation Problem”⁵², and “Ontology Tools”.

The wiki has witnessed the birth and discussion of several innovative ideas, such as “Contextual Serendipity” in the context of online systems⁵³, or “Emergent [Information] Architecture”⁵⁴, which itself builds on other ideas explored in the wiki such as “Innovation Architecture”, “Hypertext Gardens”, “Contextual Serendipity”,

⁵²A hypothetical condition in which people dropped into a rich hypertext quickly become lost and confused.

⁵³URL: <http://www.iawiki.net/ContextualSerendipity>

⁵⁴URL: <http://www.iawiki.net/EmergentArchitecture>

and even ideas from other wikis such as “Forgetfulness as Filter”⁵⁵. Overall, the contents form a quite coherent body of knowledge on information architecture.

Most of the pages in the IA wiki have been edited by several contributors. A custom has developed of leaving signatures trailing contributions, which enables credit to be attributed. There is evidence of interactions between contributors on several pages. For instance, one person might write a question on a page and another will respond to it; or someone leaves a follow-up comment below what another has written.

6.3.1.2 Weblog kitchen

The Weblog Kitchen wiki was created in the summer of 2002 by Mark Bernstein, as a place “to explore current research in weblogs, wikis, and related hypertext technologies.”

As of the end of 2002, the wiki has grown to about 250 hypertext pages. Figure 6.12 shows the contents of a page called “Weblog Theory”, as it appeared in November 2002.

Other topics under discussion include “Weblogs in research”, “Weblogs in business”, “Audience size”, “Permalinks”, “Weblog Ethics”, “Personal Voice”, “Emergent Structure”, “Document Mode” vs “Thread Mode”, “Refactoring text” vs. “Annealing Text”, and “WikiLogs”.

As of November 2002, the Kitchen has 19 “officially listed” contributors, though anonymous contributions happen once in a while. Most of the contributors are researchers. As weblogs and wikis are an emerging phenomenon whose study does not fall squarely within an existing field, contributors have a variety of backgrounds:

⁵⁵URL: <http://c2.com/cgi/wiki?ForgetfulnessAsFilter>

Weblog Theory

[Edit](#)

[Weblog Kitchen](#)
[Welcome Visitors](#)
[Things To Do](#)

[Weblog Issues](#)
[Weblog Theory](#)
[Hypertext Theory](#)
[Weblog Books](#)
[Notable Weblogs](#)

[Recent Changes](#)
[Most-Linked pages](#)
[Whos Who](#)

[Exitgate](#)
[Hypertext Kitchen](#)
[Tinderbox](#)
[Storyspace](#)

[Style Guide](#)
[How To Edit](#)
[Wiki Sandbox](#)

Scholarship about wikis, weblogs, and related phenomena.

Psychology of Weblogs A set of thought-provoking essays from online psychologist Dr. John M. Grohol about the theory and hyperbole surrounding the blogging community. [1]

Blogging Thoughts A broad, critical look at personal publishing as a research tool, by Jill Walker and Torill Mortensen. [2] (pdf)

Links and Power: The Political Economy of Linking on the Web Presented at HT02, this paper argues that links are the currency of the Web and that scholars who ignore this economy are doomed to irrelevance. [3]

I think that *traffic* is the currency of the Web, and links are the means of production. It's a small detail. This may be an important paper to the [Hypertext Conference](#) community if people hear its message. ...
[Mark Bernstein](#)

k-logs Using weblogs for personal and corporate knowledge management. Blogs provide an alternative channel to departmental and community intranet pages. When blogs are aggregated and their content 'streamed' aka blogrolling, there is the opportunity to start an informal *corporate memory*. Here are some notes. [4]

--[Denham Grey](#)

Personal knowledge publishing and its uses in research by Sébastien Paquet offers an overview of the weblogging phenomenon and a description of how it affects knowledge sharing practices, in particular in the world of research. [5]

Related [Weblog Issues](#), [Weblog Species](#)

-- Last edited November 30,
2002

[Weblog Kitchen](#) | [Contact](#)

Figure 6.12: The page “Weblog Theory” on Weblog Kitchen.

- Mark Bernstein and Tim Miles-Board are hypertext researchers;
- Jamie Blustein researches human-computer interaction;
- Les Carr researches hypertext semantics and organisation;
- Andy Edmonds has a background in cognitive psychology and does research on human factors;
- Kenneth Tompkins has a literature background;
- Jill Walker primarily researches digital narratives;
- Lauren Walker has a journalism background;
- Alex Halavais researches communication;
- Elizabeth Lane Lawley researches information science;
- Sébastien Paquet researches scholarly communication.

An interaction dynamic similar to the one described above regarding IA Wiki prevails in the Weblog Kitchen, with many pages bearing the mark of several authors.

6.3.1.3 Analysis

The two examples studied above indicate that it is possible to build a community of people with an interest in an interdisciplinary topic using an open shared knowledge repository. The fact that communities developed at all within such an information medium can be ascribed to the fact that it allowed things that were more difficult than with other communication media.

As was described above, a coherent body of knowledge has grown on both sites, which supports my second hypothesis. Furthermore, the interconnections between pages indicate that new bits of knowledge have not only been synthesised but also used and linked together.

It must be noted that the fields that constituted the central focus of both repositories are related to hypertext systems. People with an interest in those fields are typically not only familiar with hypertext as readers, but as authors as well. As a consequence, it might be easier for them to approach open shared knowledge repositories.

This is a valid observation. It would arguably be more difficult to deploy open shared knowledge repositories in populations where ability to read or write hypertext cannot be as much taken for granted. However the general popularity of wikis indicates that the skill is not very difficult to acquire, hence the barriers seem surmountable but slower adoption curves should be expected in fields unrelated to hypertext technology.

6.3.2 User survey

The second part of the survey described in Section 6.2.2 was concerned with open shared knowledge repositories. Respondents visited a web page⁵⁶ listing 15 questions. Of these, 13 were multiple-choice questions and the remaining two were open-ended. Participants filled in the questionnaire and had to click the “submit” button when done. As respondents could not be expected to be familiar with the precise meaning of the term “open shared knowledge repository”, the fuzzier term “community wiki” was used.

⁵⁶URL: <http://blogstreet.com/surveys/wiki.html>

6.3.2.1 Questionnaire

The questions that were asked follow:

1. Have you ever visited a community wiki?
2. How often do you visit wikis?
3. How often do you visit wikis to locate interesting work-related material?
4. What percentage of the time that you spend reading for professional purposes is devoted to reading material in wikis?
5. Are there wikis that you regard as being among the best sources of information on a particular topic?
6. Are there wikis that you regard as being among the best sources of knowledge on a particular interdisciplinary topic?
7. Have wikis helped you find people with professional interests close to yours but with a different professional background than yours?
8. If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through wikis?
9. Have you ever edited a wiki page?
10. How often do you edit wiki pages?
11. Based on your experience, would you say that a wiki enables the formation of a community of people with an interest in an interdisciplinary topic more easily than other available means of communication?
12. Professionally, you would describe yourself as? (a technologist, a journalist, a librarian, an educator, a lawyer, a consultant, a researcher, a knowledge worker but none of the above, not a knowledge worker)
13. How many years of professional experience do you have?
14. (Optional) In your own words, would you say that wikis have changed the way you share and obtain work-related knowledge? In what way?
15. (Optional) If you have a home page or weblog, what is its URL?

6.3.2.2 Results and analysis

The complete results of the survey are available on the World Wide Web⁵⁷.

In all, 168 people answered the survey⁵⁸. Of these, a hundred and twelve had visited a wiki at some point, and seventy-six people indicated their professional orientation. Of these 76, 45% described themselves as “a technologist”, 12% as “another kind of knowledge worker” (i.e. not one of the choices given), 17% as “a consultant”, 12% as “a researcher”; the remaining 14% were distributed relatively evenly among educators, journalists, librarians, lawyers, and people who were not knowledge workers. As regards professional experience, 21% had five years or less, 24% had from 6 to 10 years of experience, 37% had from 11 to 20 years, and 18% had more than 20 years of experience.

In this section, I present and interpret the data that is relevant to evaluating the hypotheses that were put forth at the end of Section 4.5.

Hypothesis 1: “Open shared knowledge repositories allow researchers to build a community of people with an interest in an interdisciplinary, possibly ill-defined topic.”

The responses to questions #7, #8, and #11 are relevant here. Figure 6.13 and Table 6.6 illustrate the distribution of answers to question #7, in the overall sample and in the subset of researchers (9 respondents).

In the general sample, one half of the people who answered the question indicated that they had found people with professional interests close to theirs but with a different professional background using an open shared knowledge repository. In

⁵⁷URL: <http://www.iro.umontreal.ca/~paquetse/KnowledgeSharingSurvey.html>

⁵⁸Eighty-five of them also answered the survey in Section 6.2, making for an approximate 50% overlap between the two samples.

7. Number of people with shared interests

but different background found via wikis

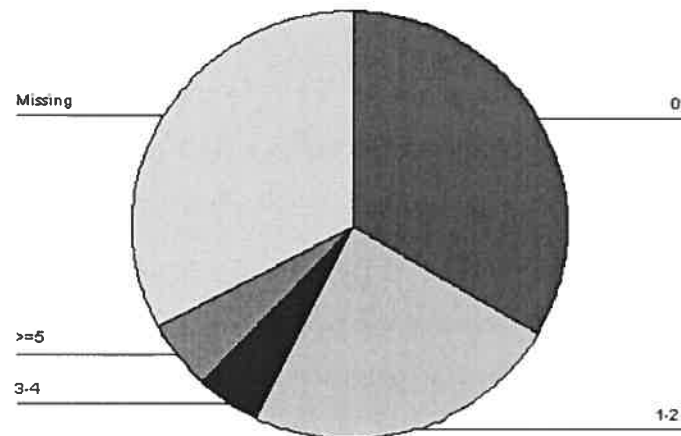


Figure 6.13: Distribution of answers to question #7, “Have wikis helped you find people with professional interests close to yours but with a different professional background than yours?”.

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, ≥ 5	9	5.4	8.0	8.0
	Yes, 3-4	8	4.8	7.1	15.0
	Yes, 1-2	40	23.8	35.4	50.4
	No	56	33.3	49.6	100.0
	Total	113	67.3	100.0	
Missing		55	32.7		
Total		168	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, ≥ 5	0	0.0	0.0	0.0
	Yes, 3-4	1	11.1	11.1	11.1
	Yes, 1-2	2	22.2	22.2	33.3
	No	6	66.7	66.7	100.0
	Total	9	100.0	100.0	
Missing		0	0.0		
Total		9			

Table 6.6: Distribution of answers to question #7, “Have wikis helped you find people with professional interests close to yours but with a different professional background than yours?”.

the researchers sample, however, this proportion is one third. This provides a limited amount of support to the hypothesis. However it does not contradict it, because it is possible that researcher respondents who answered negatively either haven't spent a lot of time navigating repositories or were not looking for such people. The average numbers are 1.46 and 1.06 contacts, respectively, for the global and the researcher sample.

Question #8, whose answer distribution is illustrated by Figure 6.14 and Table 6.7, asked if the interpersonal connections of question #7 could have been expected to be made in another way. In the general sample, 58% of respondents who gave an answer indicated that they would not, or probably not, have made such interdisciplinary connections outside of the knowledge repositories that they visited. All the researchers who made such contacts indicated that they would not, or probably not, have occurred in any other way.

This data provides moderate support to my hypothesis, as it shows that at least in some cases the usefulness of knowledge repositories for generating relationships is not outshined by other means of communication.

The next question I look at examines the ease of interdisciplinary community formation as perceived by users. Figure 6.15 and Table 6.8 show the distribution of answers to question #11, in the overall sample and in the subset of researchers.

In the general sample, among those who expressed an opinion, 78% thought of knowledge repositories as being the most adequate tool for the formation of interdisciplinary communities. Among researchers who have used knowledge repositories the proportion is 83%. These accounts, based on user's experience, provides good support for the hypothesis.

To summarise, the data collected in connection with the three questions, along with testimonies provided provides reasonable (although not overwhelming) sup-

8. Could you have connected with these people
outside of wikis?

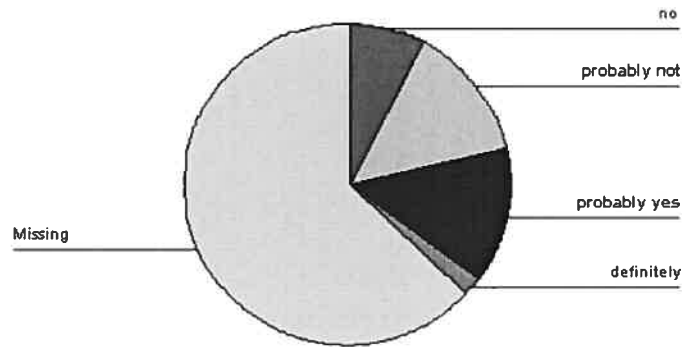


Figure 6.14: Distribution of answers to question #8, “If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through wikis?”

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	No	12	7.1	21.0	21.1
	Probably not	21	12.5	36.8	57.9
	Probably yes	21	12.5	36.8	94.7
	Definitely	3	1.8	5.3	100.0
	Total	57	33.9	100.0	
Missing		111	66.1		
Total		168	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	No	1	11.1	33.3	33.3
	Probably not	2	22.2	66.7	100.0
	Probably yes	0	0	0	
	Definitely	0	0	0	
	Total	3	33.3	100.0	
Missing		6	66.7		
Total		9	100.0		

Table 6.7: Distribution of answers to question #8, “If you answered “Yes” to the previous question, do you think you could have connected with these people in another way than through wikis?”

**11. A wiki enables an interdisciplinary community to
form more easily than other means of communication**

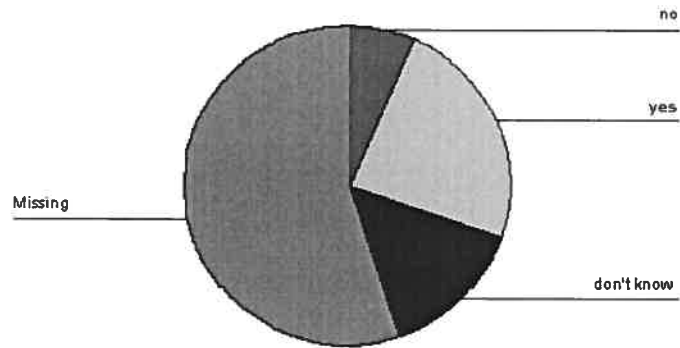


Figure 6.15: Distribution of answers to question #11, “Based on your experience, would you say that a wiki enables the formation of a community of people with an interest in an interdisciplinary topic more easily than other available means of communication?”

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	No	11	6.5	14.7	14.7
	Yes	40	23.8	53.3	68.0
	Don't know	24	14.3	32.0	100.0
	Total	75	44.6	100.0	
Missing		93	55.4		
Total		168	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	No	1	11.1	11.1	11.1
	Yes	5	55.6	55.6	66.7
	Don't know	3	33.3	33.3	100.0
	Total	9	100.0	100.0	
Missing		0			
Total		9			

Table 6.8: Distribution of answers to question #11, “Based on your experience, would you say that a wiki enables the formation of a community of people with an interest in an interdisciplinary topic more easily than other available means of communication?”

port for the first hypothesis.

Hypothesis 2: “Open shared knowledge repositories allow researchers to build a useful knowledge base around an interdisciplinary, possibly ill-defined topic.”

The responses to question #6 are relevant here. Figure 6.16 and Table 6.9 show the distribution of answers to this question.

**6. Wikis regarded as being among the best sources
of knowledge on an interdisciplinary topic**

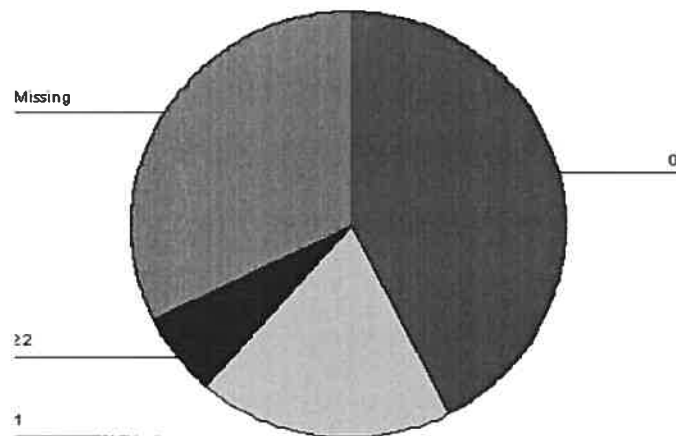


Figure 6.16: Distribution of answers to question #6, “Are there wikis that you regard as being among the best sources of knowledge on a particular interdisciplinary topic?”

In the overall sample, only 38% of respondents who were aware of knowledge repositories said they had found one or more that they considered authoritative on particular topics. Among researchers, the corresponding proportion is 33%. Although this result does not lend support to my second hypothesis, it may be explained by the current relative immaturity (and invisibility) of most existing knowledge repositories that are interdisciplinary in scope.

Examining the responses to question #14, “In your own words, would you say

<i>Global sample</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, ≥ 2	11	6.5	9.6	9.6
	Yes, one	32	19.0	28.1	37.7
	No	71	42.3	62.3	100.0
	Total	114	67.9	100.0	
Missing		54	32.1		
Total		168	100.0		

<i>Researchers</i>		Count	Percent	Valid %	Cum. %
Valid	Yes, ≥ 2	2	22.2	22.2	22.2
	Yes, one	1	11.1	11.1	33.3
	No	6	66.7	66.7	100.0
	Total	9	100.0	100.0	
Missing		0			
Total		9			

Table 6.9: Distribution of answers to question #6, “Are there wikis that you regard as being among the best sources of knowledge on a particular interdisciplinary topic?”

that wikis have changed the way you share and obtain work-related knowledge? In what way?” casts more light on this question.

One respondent indicated the superiority of wikis over message-based collective tools, indicating that he no longer bothers to read mailing lists.

“Certainly. In a few cases they have replaced usenet or mailing lists – I suppose augmented them because the news/ mailing lists still exist, but I don’t have to read them :-). The dynamics of the problem areas require consensus and consideration which are both developed over time. This isn’t possible (efficiently) in a news/mail environment where everything is ‘now’. I’ve also used them successfully over the last 3-4 years when trying to organise a lot of information (e.g. market research, software documentation, technical notes).”

Another insisted on how wikis help keep a knowledge base current, much better

than weblogs:

“Wikis foster a kind of knowledge sharing and MAINTAINABILITY that blogs really want to when they grow up some day. The updatable implicitly subject-indexed content of a wiki is wonderful for knowledge sharing. Blogs time indexed entries are a joke for that purpose.”

Yet another user was enthusiastic about using wikis as knowledge bases:

“My background is computer security, as such you learn to be pretty distrusting of humanity and generally try to lock everything down so it can't be abused. Learning that actually most people were pretty worthwhile and would contribute to an open site like [this] was revolutionary for me. I [k]now have a personal wiki, one for a community wireless group I run and one at work for the IT staff to document everything in. So, to answer your question, wiki's have made it vastly easier to compile an online, searchable and indexed knowledge base which any organisation can use.”

Finally, one user underscored the importance of wikis' ease of use.

“They allow easy dissemination of project related information. The value is in their ease of use. No other knowledge management software I have used comes close. The strange thing is that all of the wikis I have used have had excellent usability. It appears that they all spring from a common root usability model and are thus, all very friendly.”

However, a contrarian perspective was offered by one respondent (in French, translation follows):

“Les centres d’intérêt des Wiki sont trop restreints et trop ‘techniques’.”

“Wiki interest centers are too restricted and too ‘technical’.”

My own experience partially corroborates this last view, as I have encountered many wikis that had a restricted and rather technical scope⁵⁹. This echoes the pitfall identified in Section 4.6, under “Accessibility to outsiders”. However, the case studies that were presented in Section 6.3.1 above contradict this person’s assessment, as they exhibit wikis that have a less restricted and more open-ended scope. It would seem that the acuteness of this problem depends on how each user community defines the scope of discussion.

Time investment and interdisciplinary connections. For knowledge repositories it is also instructive to examine how much the ability to establish connections across disciplinary boundaries depends on the time that one invests in using the tool. I investigate this issue by examining the correlation between responses to question #4, which asked about the percentage of the time spent reading for professional purposes that was devoted to reading material in wikis, and question #7, which asked for the number of people with shared interests but with a different background that were found through wikis.

Figure 6.17 illustrates this relationship.

Respondents’ ability to establish connections across interdisciplinary boundaries does indeed seem to increase slightly with use of wikis. However, as was the case for weblogs (Figure 6.11), the linear regression does not indicate a very significant relationship ($R^2 = .23$). Again, it can be seen from the chart that several people were able to establish many connections in spite of investing relatively little time in wikis.

⁵⁹For example, Emacs wiki (URL: <http://www.emacswiki.org>)

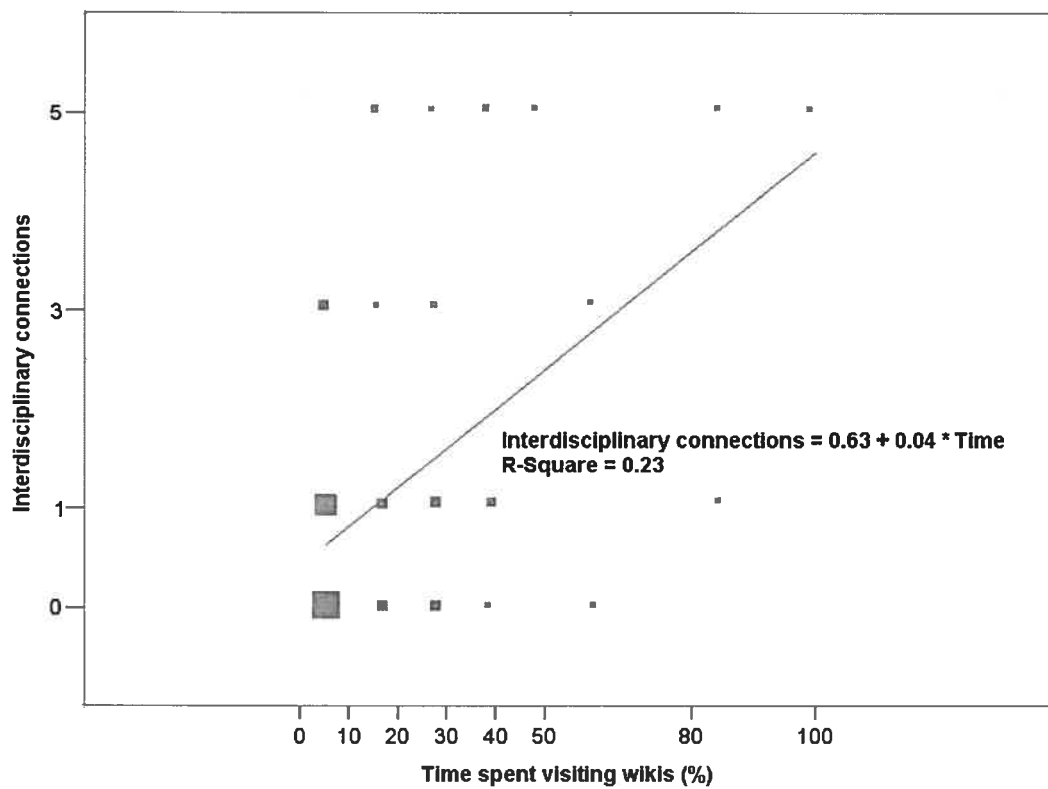


Figure 6.17: Percentage of the time spent reading for professional purposes that is devoted to visiting wikis vs. number of people with shared interests but with a different background that were found through wikis. (Square area is proportional to response count.)

Group	<i>N</i>	Mean	SD	SEM
General sample	67	.88	1.02	0.12
Researchers	9	.44	0.73	0.24
<i>t</i> value	1.235			
<i>p</i> value	0.221			

Table 6.10: Comparison of researchers and non-researchers on the number of interdisciplinary connections made through wikis.

Difference between researchers and non-researchers. Here I look for differences between the subsample of researchers and the rest of the sample, which would indicate that the findings above obtained with the general sample might not apply to researchers. Table 6.10 compares the distributions of the “number of interdisciplinary connections” variable. Based on this test, there does not appear to be significant differences between the samples.

6.3.3 Evaluation criteria

Here I review the adequacy criteria that I presented in the introduction (Section 1.1.6), and examine the degree to which Web-based open shared knowledge repositories fulfill them.

Accessibility. Wiki-type software is freely available on the Web. Installation and operation does not require much time or money. Access is limited to those researchers who can access the Web, which is a large (and growing) proportion of researchers.

Appearance of value. In both case studies above it is apparent that several people saw the value in getting involved in an open shared knowledge repository.

Adequacy. As can be judged from the survey results, a fair proportion of people who have used knowledge repositories have found it a useful medium to gather

knowledge. The results suggests however that this tool might not be suited to everybody or require an adaptation effort that is too important to make it worthwhile to some of its potential users.

Inclusiveness. This factor depends on the particular attitude of the community who uses the repository towards ideas that are foreign. Technologically, any new content can enter the repository, but whether or not it will stay there depends on where the community draws the line. So the tool supports inclusiveness but does not mandate it.

Scalability. Judging from the examples presented, it seems that open shared knowledge repositories can technically accommodate more than a hundred contributors. Socially, though, it is not obvious that such a repository will scale well. For indications on this it, is necessary to look at other repositories that are not devoted to interdisciplinary research. Two large examples exist: the original WikiWikiWeb⁶⁰ and the Wikipedia⁶¹. Both have had more than one thousand users yet remain alive and well, which is an indication of scalability.

Evolvability. Owing to their simple structure, wikis are technologically evolvable. For instance, Wikipedia successfully changed its Wiki software twice during its two years of existence. The fact that the contents may be freely reproduced and redistributed means that several different evolutionary branches can coexist, which can speed up evolution (though sometimes at the cost of creating confusion).

⁶⁰URL: <http://c2.com/cgi/wiki>

⁶¹URL: <http://wikipedia.org>

6.4 Navigable synthesis ontologies

6.4.1 Case study

In this section I report on my experience in building and using two navigable synthesis ontologies. As I was not able to find collaborators who had the time and inclination to participate in a full-fledged experiment with this tool, I worked alone in that endeavor. Consequently, the assessment that follows is much less comprehensive as that performed on the previous two tools.

I first narrate my two experiences, and follow up with observations relevant to the use of navigable synthesis ontologies for sharing knowledge across disciplines.

Seb's Web of Knowledge and Know-How. In the spring of 2001, I started building a site called Seb's Web of Knowledge and Know-How⁶², which consisted in a set of ordinary HTML pages following the hyperlinked structure of a navigable synthesis ontology. Apart from the contents not being publicly editable, the site fulfilled the criteria for being a navigable synthesis ontology.

I proceeded as follows to build this ontology: whenever I happened to come across and notice a structural similarity between concepts from different domains, I took quick notes on paper. The domains that I worked on interlinking logically included: cryptography, quantum information theory, algorithmics, computer graphics, mathematics, music, and storytelling. Now and then, I sat down to carve out appropriate definitions for each concept, keeping in mind that I needed to produce representations that exhibited the same overall basic structure.

An example of this is the link I made between the standard image rasterisation problem in computer graphics and the problem of preparing a tensor product state

⁶²<http://www.iro.umontreal.ca/~paquetse/knoweb/>

in quantum information processing, two rather specialised problems that arise in quite different fields.

In computer graphics, the *rasterisation problem* consists, starting from a three-dimensional “world” of “scene” description and a description of a camera’s parameters (such as position, orientation, and aperture angles), in generating a two-dimensional rectangular-pixel-grid image that corresponds to what the camera “sees”. In quantum physics, a *tensor product state* is a state of a composite system that is a product of (unentangled) states of each of its subsystems.

What abstraction is common to both of these problems? It may be identified by comparing some of their respective solutions. One way to rasterise a scene is to iterate over the pixels of the target image, and to compute a color value for each pixel. The computation can be done in the same way for each pixel. And one way to prepare a tensor product state is to prepare each pure state independently on the appropriate subsystem.

Both solutions are very simple, because each takes advantage of the fact that the problem it addresses can be reduced to a set of identical and independent subproblems. Thus the common abstraction that can be extracted out of this analogy is what I called an *identically separable problem* and stated as follows⁶³

- You wish to solve a problem which is decomposable into a number of identical subproblems.
- You have a way of iterating over subproblems.
- You have a unique solution pattern which is applicable to any such subproblem.

⁶³URL: http://www.iro.umontreal.ca/~paquetse/knoweb/solve_an_identically_separable_problem.html

After identifying this generic problem, I identified several other “real-world problems” that are expressions of this pattern: Eating a box of crackers, stamping a pile of envelopes, doing the dishes, cutting one’s fingernails, doing push-ups, grading a pile of multiple choice examinations, grading one multiple choice examination, etc.

I also came across problems that fit more specific patterns, such as building a car (manufacturing) or rendering a polygon list (computer graphics), which really fits the *pipelinable problem* class⁶⁴:

- You wish to solve a problem which is decomposable into a number of identical subproblems.
- You have a way of solving any given subproblem which consists in a sequence of steps (or tasks). In other words, the subproblems are separable problems.
- You have a number of agents⁶⁵, each of which is able to perform a task on a subproblem.
- You have a way for an agent to pass a problem on to another agent.

In a similar manner, I identified a number of other generic problems and was able to link them to one another and to specialisations in different areas of expertise. I also linked together various concepts that were not problems, such as musical and storytelling structures, each time explicitly defining common abstractions. In all, a little under 400 concepts were integrated.

In several cases, I had to consciously simplify the descriptions, paring them down to the essentials, in order to hide complexity and avoid having to define many

⁶⁴URL: http://www.iro.umontreal.ca/~paquetse/knoweb/solve_a_pipelinable_problem.html

⁶⁵The term “Agent” is used here in the limited sense of “Something that can perform a task when asked”.

details that were extraneous to the analogies I was formalizing. This attention often led to the identification of more basic concepts that had better be defined before the more complex ones, so as to keep definitions simple.

Generalisation/specialisation relationships enabled me to work in a manner similar to object-oriented software design, using specialisation to efficiently build descriptions of concepts that inherited properties from other concepts.

Unifying Concepts Wiki. Towards the end of 2001 I started experimenting with a Wiki-based system I called the “Unifying Concepts Wiki”⁶⁶ (UC Wiki), whose aim is “*to serve as a place to collect and organise unifying concepts, and relate them to specialised knowledge*”. Unifying concepts are understood as patterns that recur in a variety of contexts. Being able to recognise those patterns helps simplify one’s thinking and effectively solve problems. Prominent examples of unifying concepts are: evolution, symmetry, hierarchy, and self-organisation.

I briefly explain the rationale behind this statement of purpose. It is clear that the goal of connecting even a remotely significant part of specialised knowledge together cannot be achieved without a large-scale effort. However, I believe that it is possible, with modest resources, to capture a large enough number of highly useful and reusable interdisciplinary abstractions to obtain a valuable body of knowledge in the short to medium term. The prospect of helping build something that people will actually find useful can be a powerful incentive to contribute, as evidenced for instance by the success of many free / open source software and open content initiatives that rely on voluntary efforts.

Moreover, it simply makes economic sense, in terms of a usefulness-to-effort ratio, to first target the most widespread and reusable concepts. In addition, many of

⁶⁶URL: <http://www2.iro.umontreal.ca/~paquetse/cgi-bin/uc.cgi>

those concepts may provide a “backbone” that facilitates the integration of more specialised knowledge⁶⁷.

Another reason for setting this course is that unifying concepts have a kind of beauty to them that tends to attract talented individuals. “Profound”, “natural”, “simple yet powerful”, and “elegant” are adjectives often bestowed upon such concepts by those individuals. Since a project such as this needs the input of very able people, I hope that this orientation will encourage at least a few of those individuals to share their knowledge.

The site’s content is distributed under the GNU Free Documentation License⁶⁸, meaning that it can be freely copied and redistributed. Currently, most of the content gravitates around the themes of symmetry and hierarchy.

I developed the ontology much in the same way I did with the previous implementation, but took advantage of the ability to make ‘Prospective links’ to concepts not yet defined. The hypertext visualisation in my implementation highlights such holes by marking them with a question mark, which facilitates consolidation.

Several concepts that I had never before encountered in an explicit form had to be named and defined to serve as junction points in the hierarchy. Such concepts sometimes had to be inserted in between a child and its parent, but that did not pose any particular problem as it did not invalidate any other relationship.

Observations. I conclude this section with a few observations on navigable synthesis ontologies.

Building these ontologies helped me realise how much there is an abundance of tacit assumptions in the ideas that we use day-to-day. Making everything explicit

⁶⁷In the area of software development, a similar rationale underlies the development of the Standard Template Library [SL94].

⁶⁸URL: <http://www.gnu.org/copyleft/fdl.html> .

is a lot of work but it illuminates the deeper structure of our thoughts. This work had the benefit of letting me view some concepts with a more detached perspective — as it were, it gave me an “outside view” rather than an inside view on ideas that I was familiar with.

This construction endeavour also enabled me to uncover deep connections between concepts that, to my knowledge, had not been linked previously. This helped improve my understanding of both ends of these connections. It also enabled me to extract the essential structures that these concepts had in common and give them names — for instance, “Independently separable problem”. Having isolated and identified such fundamental structures enabled me to recognise them elsewhere more readily. I realised that, much as a modern-day Mr. Jourdain, I’d been using some of these structures most all my life without recognizing them!

On a few occasions, I have gone back to the ontologies several months after I had stopped developing it, to look for concepts whose name and definitions I did not recall clearly. Each time, I started at a point in the ontology that had something to do with what I was looking for, then used the definitions and links in addition to my judgment to navigate efficiently towards my goal. I give two examples here.

In the first case, I recalled having entered a concept into the first ontology that referred to artifacts that could be divided into well-delineated parts. I knew that what I called “chorus-verse songs” were a specific case of that concept. A search led me directly to that concept; from there I generalised several times, until I reached the “Block-Structured Artifact” concept which was what I had been looking for. From there I explored some of the various specialisations to refresh my memory.

In the second case, I was looking for a particular approximate root-finding algorithm similar to Newton’s method but that did not require computing the derivative of the input function.

I seemed to recall that both algorithms had a certain amount of similarity with the binary search algorithm, and so started off by searching for “binary search” in UC Wiki. I found the page describing that algorithm⁶⁹ (shown in Figure 6.18). That page was linked to another describing the problem it solved, namely, searching in a one-dimensional, finite, ordered set⁷⁰; this is where I went next.

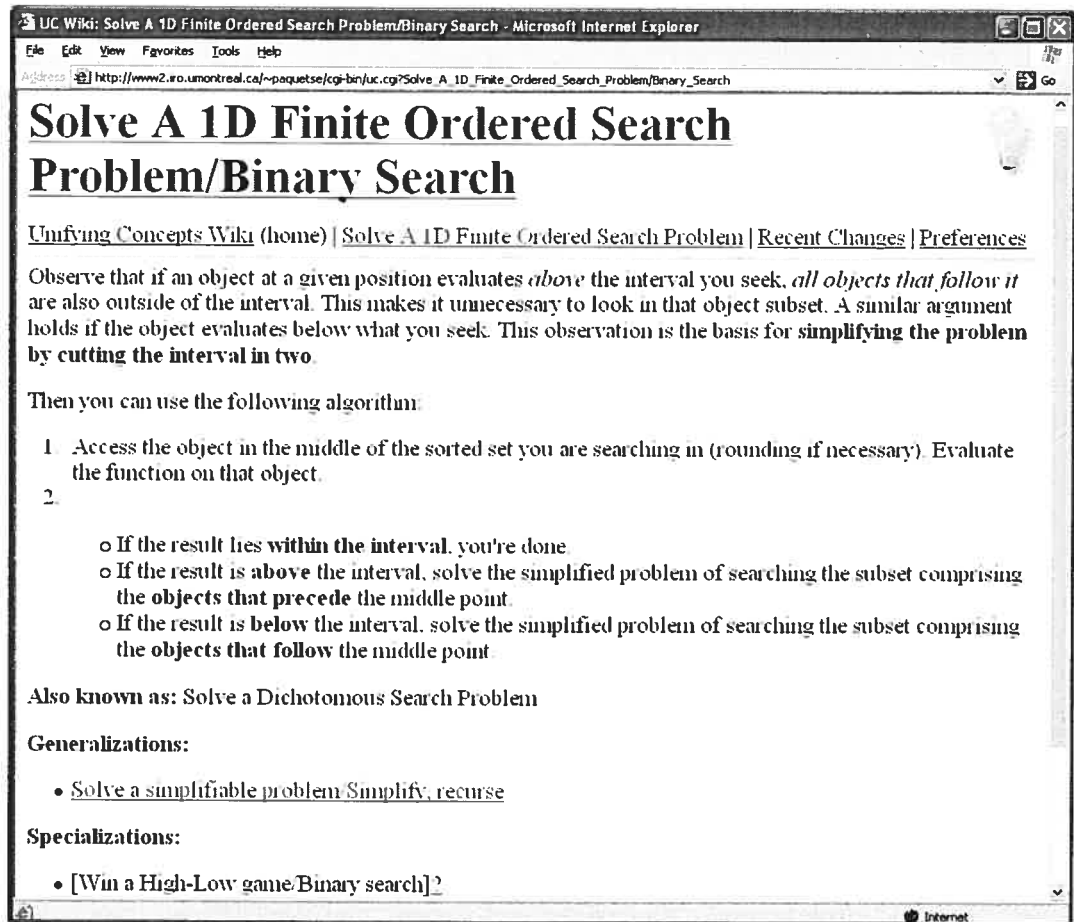


Figure 6.18: Part of the page for the “Binary Search” concept in UC Wiki.

I reasoned that, in the case of a continuous function, I had to abstract out the finiteness and order properties, and followed generalisation links to work my way

⁶⁹URL: http://www2.iro.umontreal.ca/~paquetse/cgi-bin/uc.cgi?Solve_A_1D_Finite_Ordered_Search_Problem/Binary_Search

⁷⁰URL: http://www2.iro.umontreal.ca/~paquetse/cgi-bin/uc.cgi?Solve_A_1D_Finite_Ordered_Search_Problem

up to the concept called “Solve a 1D Bounded Search Problem”⁷¹. There I found a link to the specialisation called “Approximate a root of a continuous real-valued function of a real variable, given bounds and a means of evaluation”, which I recognised as the problem solved by the algorithm I was seeking. I followed the link and drilled down to the problem “Approximate a root of a continuous real-valued function of a real variable, given opposite-valued bounds and a means of evaluation”, which linked to a solution algorithm called “Bisection method”⁷². I knew I had reached my destination.

Now, one could imagine a situation in which a programmer with little (or rusty) knowledge of numerical methods, but who is familiar with binary search in sorted tables, has at some point to solve a complicated equation approximately. If he had the intuition that a method similar to binary search might help, the programmer could follow essentially the same path I did to end up at the bisection method.

Conclusion. Overall, my observations offer support to the hypotheses I stated in Section 5.7, to the effect that 1) analogies are powerful ways to make knowledge accessible across disciplines; and that 2) navigable synthesis ontologies enable analogies to be organised in such a manner that finding concepts analogous to a given concept becomes easy. However, experimentation with a group of users is obviously necessary to further establish the validity of that claim.

⁷¹URL: http://www2.iro.umontreal.ca/~paquetse/cgi-bin/uc.cgi?Solve_A_1D_Bounded_Search_Problem

⁷²URL: http://www2.iro.umontreal.ca/~paquetse/cgi-bin/uc.cgi?Approximate_A_Root_Of_A_CRVFOARV,_Given_Opposite-Valued_Bounds_And_A_Means_Of_Evaluation/Bisection_Method

6.4.2 Evaluation criteria

Let us now examine navigable synthesis ontologies in light of the characteristics that were stated as desirable in the introduction.

Accessibility. As my implementation demonstrates, it is possible to implement such an ontology on the World Wide Web using tools that are available to everyone who is connected to the Internet.

Appearance of value. As this tool has not been widely publicised yet, it is difficult to see whether it appears to be of value for researchers, although the reader by now has probably formed an opinion by himself.

Adequacy. I have argued that this tool enables people to find relevant knowledge that would have been practically impossible to find otherwise. This provides an incentive for *using* the tool, but not for *building* its contents. I surmise that the building activity primarily answers the needs of teachers, evangelists and the like, who are keenly interested in spreading ideas they find valuable. In a realistic scenario, there may thus be many more users than builders of navigable synthesis ontologies (just as there are many more students than teachers, and vastly more users than authors of dictionaries).

Inclusiveness. An ontology such as I have described allows any kind of knowledge to be represented without discrimination, so long as the representation is logically coherent.

Scalability. This characteristic depends on the particular implementation. Our implementation is based upon a centralised wiki. Wikipedia, the most active wiki, allowed several hundred collaborators per day to work together and create over 100,000 pages. This indicates that a centralised wiki-based design scales at least up to this size. However, it tells nothing about larger scales.

Evolvability. The system allows for a change in format by virtue of having a consistent structure across pages. However, barring the development of automated natural language understanding, a migration of the content to an equivalent representation in a completely formal knowledge representation language (such as the Knowledge Interchange Format, KIF [GF92]), or to representations other than hypertext, would in all likelihood require human involvement.

Chapter 7

Conclusion

We are at the very beginning of time for the human race.

It is not unreasonable that we grapple with problems. But there are tens of thousands of years in the future. Our responsibility is to do what we can, learn what we can, improve the solutions, and pass them on.

Richard Feynman

In this chapter I recapitulate the contents of the thesis, compare the tools that I have studied, and provide directions for future work on interdisciplinary knowledge sharing.

In this thesis, I have described the problem of interdisciplinary knowledge sharing as the increasingly pronounced inability of specialists to locate and/or make proper use of the results and lines of thought developed in disciplines other than their own, even when the problems they are tackling are substantially similar.

This situation arises as a consequence of the development of increasingly specialised languages by groups with a common interest, which results in more ef-

ficient communication amongst members at the price of introducing a learning phase where newcomers must assimilate the concepts and the culture of the group.

The interdisciplinary language barrier problem has several negative consequences. First, it hinders work on the increasing number of important interdisciplinary problems that arise in research, by making it difficult for people with different backgrounds to understand one another. Second, failure to identify ideas that are related across disciplines results in frequent reinventions of the wheel. Redundant material is piled up on top of a mountain of publications, exacerbating an already painful information overload situation among researchers. Third, the lack of communication between fields makes it harder for a field to assess and explain its relevance to the evolution of knowledge and its contribution to other fields and to society in general. As knowledge grows and splits into more and more subfields, the specialised researcher commands a steadily decreasing proportion of the total knowledge available and is less and less apt to form a global perspective in which to align his work.

I have pointed out that at its root this problem combines two challenging issues, one social and the other technological in nature. The social aspect relates to the fact that scholars have frequent contacts with people within their primary research community and comparatively infrequent contacts with people outside that community. As a result, they are not aware of “who knows what” outside their field, which makes it difficult to find knowledge. Moreover, they cannot take advantage of the serendipitous transmission of knowledge that occurs when people with different backgrounds are in somewhat regular contact.

From a technological standpoint, the missing links are better interfaces to knowledgeable people and communities, and directly to knowledge in the form of information resources. When someone comes up with an interesting, seemingly new idea in which would like to dig, it should be possible for them to tap into a net-

work of people with diversified interests, in order to quickly home in on people doing related work or on related textual material. If they should find themselves in contact with a few people sharing an interest, it should also be possible for them to build a community of knowledge around that idea, regardless of whether the idea fits well or not within an already established field. Ideally, it should also be possible for a researcher to locate ideas related to his own by navigating a knowledge base by himself, without having to solicit help from colleagues.

A well-developed tool for the last task does not exist yet. Traditional means exist for achieving the first two of the tasks I have just listed, but they are insufficient. Scholarly networking requires significant investment and the nature of the established interpersonal links still depends on chance as much as on genuine intellectual affinity, and interdisciplinary networking is heavily constrained by the difficulty of organizing successful social events that bring together people with different backgrounds. Building communities of knowledge the old-fashioned way typically takes years of commitment, a string of focused, visible publications, and enough funding to organise gatherings. So new, more powerful tools are desirable.

In this thesis, I have proposed to apply three tools to the problem of facilitating knowledge sharing across disciplines: personal knowledge publishing, open shared knowledge repositories, and navigable synthesis ontologies. I have also assessed the usefulness of these three tools toward that goal.

Personal knowledge publishing / weblogs. The first tool, personal knowledge publishing, is an activity where individual researchers regularly update a website whose content consists in individual posts that feature hyperlinks and are presented in reverse chronological order. The website is public and its contents are freely accessible on the World Wide Web. Archives of old posts are kept in such a way that links to them (be they internal or external) still work after they

have left the front page.

Personal knowledge publishing fosters interdisciplinary knowledge sharing in several ways. First, acting as a radically open conversational medium with low barriers to entry, it allows ideas to be exchanged and feedback to be obtained easily in a manner that is unconstrained by disciplinary boundaries, even when those ideas are not fully worked out or break with tradition.

Second, acting as an information filter, it enables content relating to emerging interdisciplinary fields to be clustered and made visible much more efficiently than the traditional publication system allows.

Third, it enables a web of trust to be built that is again unconstrained by disciplinary boundaries. Since the frequent-update characteristic is prone to induce repeat visits and conversations, persistent interpersonal trust relationships can be established and gradually strengthened among researchers. This enables the development of a new, practically noise-free, social network where the linkage depends more on intellectual affinity than on membership in the same research community.

Fourth, the network can serve to efficiently assemble groups around emerging interdisciplinary ideas, because it implements a very efficient idea dissemination system. When a promising idea comes up, the word spreads quickly through the network and in a matter of days it is possible for people who do not know one another but happen to share a common interest to find one another.

Three case studies and a survey conducted with users of that tool have confirmed the usefulness of this tool with respect to the goal of interdisciplinary knowledge sharing.

Open shared knowledge repositories / wikis. The second tool, the open shared knowledge repository, is a collective space made up of pages and links between pages. Its contents are wholly accessible and managed by an open community and represent the community's agreed-upon knowledge. The content of the knowledge repository is freely accessible and may be freely reproduced or redistributed.

Open shared knowledge repositories foster interdisciplinary knowledge sharing in two ways. First, they enable people with a common interest in an issue that does not fit neatly within a particular discipline to collaboratively create, share and structure their knowledge, and to relate it to established knowledge. Second, they enable them to work out ways of explaining their findings to a wider audience and of giving visibility to the issue under discussion. This can attract more collaborators from a diverse basin and potentially help the community reach the critical mass necessary for getting recognition and giving an air of legitimacy to its ideas.

An open shared knowledge repository can work in synergy with a personal knowledge publishing network; conversations are carried out on the network and agreed-upon knowledge is accumulated in the repository.

Two case studies and a survey conducted with users of that tool have also confirmed the practical usefulness of this tool with respect to the goal of interdisciplinary knowledge sharing.

Navigable synthesis ontologies. The third tool, the navigable synthesis ontology, is a specialisation of an open shared knowledge repository in which each page gives one or more complete formal definitions of a concept, referencing other concept pages as needed, and features links to generalisations and specialisations of the concept. Each page may point to a subpage that lists things that can be considered as instances of the concept, and to a subpage with links to external

documents that specifically discuss the concept under consideration. In addition, each page may point to a subpage for discussion of its contents.

The “synthesis” aspect in the denomination of this object arises from its ability to link related concepts from different areas together by finding unifying concepts that capture the commonality.

A navigable synthesis ontology enables its users to navigate knowledge in a logical way and reach sought-after concepts regardless of whether one knows their name or not. It can be grown organically through the participation of many contributors. It can also provide interpersonal networking benefits similar to those afforded by an open shared knowledge repository.

A case study has supported the practical usefulness of this tool for connecting knowledge across fields, but more extended experimentation is necessary to confirm that it is effectively usable by a community of researchers.

Comparison of the tools and their uses. All three of the tools that I have studied are useful towards interdisciplinary knowledge sharing, but each has its particular strengths and appropriate contexts of use. Table 7.1 summarises a comparison of the tools along several dimensions.

The first dimension relates to the users. Personal knowledge publishing is suited for a single user, while open shared knowledge repositories are intended for community use. Navigable synthesis ontologies can be used by a single person to draw connections across disciplines, but its power comes from collective use.

The second dimension is the form that knowledge assumes in the tool. In personal publishing, knowledge is structured chronologically, while knowledge in a repository is an informal semantic network in which elements are more densely

<i>Tool</i>	Personal knowledge publishing	Open shared knowledge repository	Navigable synthesis ontology
<i>Users</i>	Single user	Multiple users	Single/Multiple
<i>Structure</i>	Chronological	Freeform semantic network	Logical ontology
<i>Core Use</i>	Personal connections	Community building knowledge pooling	Knowledge structuring and organisation
<i>Dynamics</i>	New content pushed on top of old content	Everything can change and get updated	Everything can grow monotonically
<i>Approachability</i>	Easy (personal, informal)	Medium (social, informal)	Hard (formal thinking)

Table 7.1: Comparison of the tools.

interlinked. In a navigable synthesis ontology, knowledge is represented by a logically structured ontology.

The third dimension relates to the principal way in which each tool facilitates interdisciplinary knowledge sharing. Personal knowledge publishing does so mainly by linking people with common interests together across disciplinary barriers, creating a network of individuals. Open shared knowledge repositories let people create communities and pool together knowledge that is of interest to the community. Navigable synthesis ontologies help organise knowledge to make it unambiguous and more easily retrievable.

The fourth dimension is the dynamics of use associated with each tool. Personal knowledge publishing works by “pushing” recent content on top of a stack of older content. Older content is not modified and may eventually become outdated and fall into disuse. Open shared knowledge repositories work differently: everything is kept up to date, content is continually refreshed and expands in several directions at once. Synthesis ontologies define increasingly rich models that do not need to be changed but may be replaced with more useful ones.

The fifth dimension is approachability, i.e. how easy it is to understand and begin

using the tool. Personal knowledge publishing is the easiest to use, with most people who are acquainted with the World Wide Web quickly grasping how it works and what it is about. Open shared knowledge repositories are more tricky to understand, among other things because they involve making up new names, explicitly interlinking ideas, and negotiating the knowledge creation activity with other people. The navigable synthesis ontology is the hardest form of all, as it involves the difficulties in knowledge repositories but additionally requires careful logical thinking and the ability to synthesise generalisations.

Consequently, people who wish to engage in interdisciplinary knowledge sharing using these tools would do well to start out with personal knowledge publishing, then move on to open shared knowledge repositories as their social network develops and a shared focus emerges. A navigable synthesis ontology should come last (if at all), as it requires a significant investment and has a very long-term payoff.

The tools have several characteristics in common as well. First, they build on hypertext in a fundamental way and are consequently much more efficient if the user is familiar with this technology, as an author as well as a reader. Second, because they all use the Internet, they diminish barriers to collaboration across physical distance.

Third, actively producing material with any of these tools is obviously undesirable for people who would rather keep their ideas to themselves. Using them requires a commitment to early (i.e., pre-review) public sharing of knowledge, and a willingness to risk making public mistakes. It must however be kept in mind that this is not an all-or-nothing proposition: it is possible to share some, but not all, of one's ideas. Unlike most other computer tools, the benefits that come out of use are not guaranteed because they depend on other people. A certain amount of faith in reciprocity and cooperation is thus necessary.

And last but not least, all of the tools present the most value to people who do

not yet have a well-developed social network that they can tap into for reliable knowledge. For this reason, it can be expected that they will mostly be adopted at first by students and young scholars. There is however reason to hope that in the long term, more established researchers will find these media useful as well.

Future work. There are several promising directions for future work into interdisciplinary knowledge sharing that extend beyond what I have presented in this thesis.

First, the development of smart navigators, that is, tools that take a user's profile and lead him to very relevant sites in a personal knowledge publishing network or to very relevant knowledge repository pages, is desirable because exploration can be time-consuming, especially for new users.

Second, conversations held through personal knowledge publishing are often difficult to track. Appropriate standards for conversation mapping, and the development of conversation visualisation tools, would make personal knowledge publishing an even more compelling communication tool.

Third, it is desirable to better bridge personal knowledge publishing and shared knowledge repositories. Currently, personal knowledge publishing facilitates one-to-one networking but offers little explicit support for clustering of people into groups.

One promising idea is to combine aspects of both technologies by creating *blog channels*. A blog channel is a topical weblog to which all weblog editors can contribute posts from their own weblog. Blog channels implicitly create loosely joined interest groups, which may eventually go on to creating an open shared knowledge repository if enough cohesion is found between contributors to the channel. I have already undertaken an investigation of that research direction by designing and

implementing the Internet Topic Exchange¹ together with Phillip Pearson[PP03]. Improving and evaluating this resource is part of my future research plans.

Another connection between personal knowledge publishing and shared repositories that could be worth exploring is to use personal knowledge publishing as a way to reliably assess authorship of contributions to shared repositories. This would at the same time serve to unify individuals' identities in the various shared repositories they might participate in.

Fourth, introducing the array of tools to already formed research communities that bring together people with diverse backgrounds could be an interesting avenue to experiment. Together with a few other researchers, I have recently proposed an initiative along these lines aiming to introduce personal knowledge publishing and a shared repository in the quantum information processing community[ABMP02].

Finally, as regards navigable synthesis ontologies, two improvements could make the user experience more powerful. The first would consist in collecting data on how much each link is traversed to dynamically influence the way those links are displayed on each page. This could for instance enable visitors to see which paths are more well-travelled than others.

Another enhancement would be to use personal user profiles (as provided for example by the content of a personal publication) to influence the display of links for each individual user, leading him to pathways through the knowledge space that stand more chances of being familiar.

As my assessment of navigable synthesis ontologies was incomplete, it will be necessary to experiment with that tool on a larger scale to better assess their usefulness for interdisciplinary knowledge sharing.

¹URL: <http://topicexchange.com>

The evolution of all tools will obviously benefit greatly from the incorporation of as much user input as possible.

Appendix A

Cast of Characters

The table in the next page lists the people who are mentioned at one point or another in Chapter 6 of this thesis and their affiliation as of January 2003, to the best of my knowledge.

Name	Position	Affiliation	Country
Mark Bernstein	Chief scientist	Eastgate Systems	USA
Jamie Blustein	Professor	Dalhousie University	Canada
VeerChand Bothra	unknown	unknown	India
Les Carr	Lecturer	University of Southampton	United Kingdom
Stephen Downes	Researcher	National Research Council	Canada
Andy Edmonds	Ph.D. candidate	Clemson University	USA
Lilia Efimova	Ph.D. candidate	Telematica Institut	The Netherlands
Sebastian Fiedler	Ph.D. candidate	Universität Augsburg	Germany
Alison Fish	unknown	unknown	USA
Jim Flowers	Professor	Ball State University	USA
Denham Grey	Consultant	Greymatter	USA
Kat Hagedorn	Metadata Librarian	University of Michigan	USA
Eric Hanson	unknown	unknown	USA
Alexander Halavais	Professor	Buffalo University	USA
Spike Hall	Professor	Drake University	USA
Pete Harbeson	Technical Writer	unknown	USA
Elizabeth Lane Lawley	Professor	Rochester Inst. of Technology	USA
Victor Lombardi	Information architect	Razorfish	USA
Ron Lusk	unknown	unknown	unknown
Al Macintyre	Programmer	Global Wire Technologies	USA
Jim McGee	Professor	Kellogg School of Management	USA
Tim Miles-Board	Researcher	University of Southampton	United Kingdom
Peter Morville	President	Semantic Studios	USA
Matt Mower	Consultant	Novissio	United Kingdom
Tom Munnecke	Visiting Scholar	Stanford University	USA
Sébastien Paquet	Ph.D. candidate	Université de Montréal	Canada
Phillip Pearson	Software developer	Myelin Electronics	New Zealand
Tanya Rabourn	Consultant	Academic Information Systems	USA
John Robb	CEO	Userland	USA
Louis Rosenfeld	Consultant	Louis Rosenfeld LLC	USA
Seth Russell	unknown	unknown	unknown
Greg Searle	Chief Technology Officer	Tomoye	Canada
Rashmi Sinha	Lecturer	UC Berkeley	USA
Dolores Tam	Translator	unknown	Canada
Roland Tanglao	unknown	unknown	Canada
Kenneth Tompkins	Teacher	Richard Stockton College	USA
Laura Trippi	Professor	Simon Fraser University	Canada
Jill Walker	Ph.D. candidate	University of Bergen.	Norway
Lauren Walker	unknown	Eastgate Systems	USA
Christina Wodtke	Interaction designer	Yahoo!	USA
Phil Wolff	Chief Technology Officer	Evanwolf group	USA
Ton Zijlstra	Masters student	University of Twente	The Netherlands

Appendix B

Glossary

Blog: The diminutive form of “weblog”.

Blogosphere: The growing portion of the World Wide Web that is occupied by weblogs.

Blogroll: A list of weblogs that are deemed interesting by a particular weblog editor, usually appearing in a column alongside the weblog posts.

Community weblog: A weblog that has several editors. Well-known examples include Slashdot (<http://slashdot.org>), Kuro5hin (<http://kuro5hin.org>), and MetaFilter (<http://metafilter.com>).

K-Log, knowledge log: A weblog that is used by a knowledge worker to share ideas and links, either with the world or only with the people in his organisation.

Navigable synthesis ontology: An open shared knowledge repository that is used to specify an ontology, as described in Section 5.2.

Ontology: An explicit specification of a conceptualisation in the form of a collection of named concepts, instances, relations and axioms.

Open shared knowledge repository: A website that serves to share knowledge among a community. A more precise definition is proposed in Section 4.2.

Personal knowledge publishing: A weblog where a knowledge worker or researcher makes his observations, ideas, insights, interrogations, and reactions to others' writing publicly available in the form of a weblog. Equivalently, a K-log that is available for free on the World Wide Web.

Weblog: Generally speaking, an online personal journal. There is only rough consensus on what does and what does not properly constitute a weblog. A more precise definition is proposed in Section 3.2 of this thesis.

Wiki: Generally speaking, a collectively editable website. There is not a general consensus on what does and what does not properly constitute a wiki.

Bibliography

- [ABMP02] E. Aïmeur, G. Brassard, D. Mayers, and S. Paquet. Interactive Web support for QIP.
<<http://www.msri.org/ext/InteractiveWebSupportForQIP.htm>>, accessed Nov. 29, 2002.
- [AGPLP98] J. Arpírez, A. Gómez-Pérez, A. Lozano, and S. Pinto. (Onto)²Agent: An ontology-based WWW broker to select ontologies. In *Proc. Workshop on Applications of Ontologies and Problem-Solving Methods, ECAI-98*, pages 16–24. ECAI, 1998.
- [Aro02] L. Aronsson. Operation of a large scale, general purpose wiki website. In João Álvaro Carvalho, Arved Hübler, and Ana Alice Baptista, editors, *Proceedings of the 6th International ICCV/IFIP Conference on Electronic Publishing held in Karlovy Vary, Czech Republic*, November 2002.
- [BA99] Albert-László Barabási and Réka Albert. Mean-field theory for scale-free random networks. *Physica A*, 272:173–187, 1999.
- [BA00] Albert-László Barabási and Réka Albert. Scale-free characteristics of random networks: The topology of the world wide web. *Physica A*, 281:68–77, 2000.
- [BB01] Albert-László Barabási and Ginestra Bianconi. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(4):436–442, 2001.
- [BBP97] D.H. Bailey, P.B. Borwein, and S. Plouffe. On the rapid computation of various polylogarithmic constants. *Mathematics of Computation*, April 1997.
- [BCMNO2] M. Bonifacio, R. Cuel, G. Mameli, and M. Nori. A peer-to-peer architecture for distributed knowledge management. In *Proceedings of 3rd International Symposium on Multi-Agent Systems*,

Large Complex Systems, and E-Businesses (MALCEB'2002), Erfurt/Thuringia, Germany, October 2002.

- [BD91] J.S. Brown and P. Duguid. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. Technical report, The Institute of Management Sciences (now INFORMS), 1991.
- [BD00] J.S. Brown and P. Duguid. *The Social Life of Information*. Harvard Business School Press, 2000.
- [Bec97] T. Beckman. A methodology for knowledge management. In *International Association of Science and Technology for Development (IASTED) AI and Soft Computing Conference*, 1997.
- [BF98] V.R. Benjamins and D. Fensel. The ontological engineering initiative (KA)². In N. Guarino, editor, *Formal Ontology in Information Systems*, pages 287–301. IOS Press, 1998.
- [BFDGP99] V. R. Benjamins, D. Fensel, S. Decker, and A. Gomez-Perez. (KA)²: building ontologies for the internet: a mid-term report. *International Journal of Human-Computer Studies*, 51(3):687–712, 1999.
- [BFGP98] V.R. Benjamins, D. Fensel, and A. Gomez-Perez. Knowledge management through ontologies. In *Proc. 2nd International Conference on Practical Aspects of Knowledge Management (PAKM'98)*, Basel, Switzerland, 1998.
- [BFGPGP98] M. Blazquez, M. Fernandez, J. M. Garcia-Pinar, and A. Gomez-Perez. Building ontologies at the knowledge level using the ontology design environment. In B. Gaines and M. Musen, editors, *Proceedings of the Eleventh Workshop on Knowledge Acquisition, Modeling and Management, KAW '98*, 1998.
- [BJR⁺02] A.-L. Barabási, H. Jeong, R. Ravasz, Z. Néda, T. Vicsek, and A. Schubert. On the topology of the scientific collaboration networks. *Physica A*, 311:590–614, 2002.
- [BLHL01] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 234(5):34–43, 2001.
- [Bow02] J. Bowes. Building online communities for professional networks. In *Global Summit of Online Knowledge Networks*, March 2002.
- [CN93] P. Curtis and D.A. Nichols. Muds grow up: Social virtual reality in the real world. In *Proceedings of the Third International Conference on Cyberspace*, 1993.

- [Col02a] A. Coleman. Interdisciplinarity: The road ahead for education in digital libraries. *D-Lib Magazine*, 8(7/8), 2002.
- [Col02b] IAWiki Collective. Content silos.
<<http://iawiki.net/ContentSilos>>, accessed Dec. 20, 2002.
- [Col02c] MeatBall Wiki Collective. Wittenberg door.
<<http://www.usemod.com/cgi-bin/mb.pl?WittenbergDoor>>, accessed Dec. 21, 2002.
- [Cro82] B. Cronin. Progress in documentation: invisible colleges and information transfer, a review and commentary with particular reference to the social sciences. *Journal of Documentation*, 38, 1982.
- [CS92] D. Crane and H. Small. *American Sociology since the Seventies: The Emerging Crisis in the Discipline*, chapter "Sociology and its Publics: The Forms and Fates of Disciplinary Organization". University of Chicago Press, 1992.
- [Cur02] C. Curran. The medical journal meets the Internet. *First Monday*, 7(6), 2002.
- [CW93] M.F. Cohen and J.R. Wallace. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, 1993.
- [DCGR99] R. Dieng, O. Corby, A. Giboin, and M. Ribiere. Methods and tools for corporate knowledge management. *International Journal of Human-Computer Studies*, 51(3):567–598, 1999.
- [Dew33] J. Dewey. *How We Think: A Restatement of the Relation of Reflective Thinking to the Educational Process*. Heath, 1933.
- [DGD⁺98] P. F. Drucker, D. Garvin, L. Dorothy, S. Susan, J.S. Brown, and D. Leonard. *Harvard Business Review on Knowledge Management*. Harvard Business Review Series, 1998.
- [DGKN00] J. Dealy, A. Golubowski, C. Kapa, and R. Noble. The journal crisis, information technology, and the restructuring of scholarly communication. *BiblioFile: The Libraries' Newsletter for Faculty*, February 2000.
- [Doc02] C. Doctorow. How I learned to stop worrying and love the panopticon. The O'Reilly Network,
<http://www.oreillynet.com/lpt/a/network/2002/03/08/cory_google.html>, accessed Nov. 5, 2002.

- [Dom98] J. Domingue. Tadzebao and webonto: Discussing, browsing, editing ontologies on the web. In B. Gaines and M. Musen, editors, *Proc. 11th Workshop on Knowledge Acquisition, Modeling and Management, KAW'98*, 1998.
- [Dre87] K.E. Drexler. Hypertext publishing and the evolution of knowledge. *Social Intelligence*, 1(2):87–120, 1987.
- [Dys02] E. Dyson. Parallel channel.
<<http://www.edventure.com/conversation/article.cfm?Counter=8648145>>, accessed Dec. 17, 2002.
- [Ens01] M. Enserink. Scientific publishing: Peer review and quality: A dubious connection? *Science*, 293:2187, 2001.
- [Esp01] D.P. Esparza. Therapeutic metaphors and clinical hypnosis. *Hypno-Genesis*, 2001.
- [Euz96] J. Euzenat. Corporate memory through cooperative creation of knowledge bases and hyper-documents. In *Proc. KAW '96*, 1996.
- [FFR97] A. Farquhar, R. Fikes, and J. Rice. The ontolingua server: A tool for collaborative ontology construction. *International Journal of Human-Computer Studies*, 46(6):707–727, 1997.
- [FPJ97] M. Fernandez, A. Gomez Perez, and N. Juristo. Methontology: From ontological art towards ontological engineering. In *Spring Symposium Series on Ontological Engineering*, pages 33–40. AAAI Press, 1997.
- [GF92] M. R. Genesereth and R. E. Fikes. Knowledge interchange format, version 3.0, reference manual. Technical Report Logic-92-1, Computer Science Dept., Stanford University, 1992.
- [GJ93] D. Gentner and M. Jeziorski. *Metaphor and Thought*, chapter 20, pages 447–480. Cambridge University Press, 1993.
- [GJ99] F. Godlee and T. Jefferson, editors. *Peer Review in Health Sciences*. BMJ Books, 1999.
- [Gra73] M. Granovetter. Strength of weak ties. *American Journal of Sociology*, 78, 1973.
- [Gre94] J.L. Gresham. From invisible college to cyberspace college: Computer conferencing and the transformation of informal scholarly communication networks. *Interpersonal Computing and Technology*, 2(4), 1994.

- [Gru93a] T.R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands, 1993. Kluwer Academic Publishers.
- [Gru93b] T.R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.
- [GTGB84] C.M. Goral, K.E. Torrance, D.P. Greenberg, and B. Battaille. Modelling the interaction of light between diffuse surfaces. In *Computer Graphics (SIGGRAPH '84 Proceedings)*, volume 18, pages 212–222, July 1984.
- [Hil02] J. Hiler. Google loves blogs: How weblogs influence a billion google searches a week.
<<http://www.microcontentnews.com/articles/googleblogs.htm>>, accessed Dec. 21, 2002.
- [Hil03] J. Hiler. The microcontent news glossary.
<<http://www.microcontentnews.com/resources/glossary>>, accessed Jan. 23, 2003.
- [HK02] P.M. Hildreth and C. Kimble. The duality of knowledge. *Information Research*, 8(1), 2002.
- [HKW98] P.M. Hildreth, C. Kimble, and P. Wright. Computer mediated communications and international communities of practice. In *Proceedings of Ethicomp'98*, 1998.
- [HKW00] P.M. Hildreth, C. Kimble, and P. Wright. Communities of practice in the distributed international environment. *Journal of Knowledge Management*, 4(1), 2000.
- [HWB⁺02] S. Hitchcock, A. Woukeu, T. Brody, L. Carr, W. Hall, and S. Harnad. Evaluating citebase, an open access web-based citation-ranked search and impact discovery service. Technical report, Department of Electronics and Computer Science, University of Southampton, 2002.
- [IR] Ipsos-Reid. The little Internet fact book. <http://www.ipsos-reid.com/us/services/dsp_little_net_book.cfm>, accessed Jan. 21.
- [Jar30] V. Jarník. O jistém problemu minimaalnim. *Moravske Prirodovedecké Společnosti*, 6, 1930.

- [KDM⁺01] Y. Kalfoglou, J. Domingue, E. Motta, M. Vargas-Vera, and S.B. Shum. myPlanet: An ontology-driven webbased personalized news service. In *Proceedings of the IJCAI'01 Workshop on Ontologies and Information Sharing*, Seattle, WA, USA, 2001.
- [KFK02] R. Kling, J. Fortuna, and A. King. The real stakes of virtual publishing: The transformation of E-Biomed into PubMed Central. Technical report, Center for Social Informatics, Indiana University, 2002.
- [Kle98] J.T. Klein. Notes toward a social epistemology of transdisciplinarity. *Bulletin Interactif du Centre International de Recherches et Études transdisciplinaires*, 12, February 1998.
- [KLW95] M. Kifer, G. Lausen, and J. Wu. Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 42(4):741–843, 1995.
- [KMG95] P.D. Karp, K.L. Myers, and T. Gruber. The generic frame protocol. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence (IJCAI'95)*, 1995.
- [KSS97] H. Kautz, B. Selman, and M. Shah. The hidden web. *The AI Magazine*, 18(2), 1997.
- [Lak87] G. Lakoff. *Women, Fire, and Dangerous Things*. University of Chicago Press, 1987.
- [Lan00] R.N. Langlois. Modularity in technology and organizations. *Research Papers in Network Institutional Theory*, February 2000.
- [Law01] S. Lawrence. Online or invisible? *Nature*, 411(6837):521, 2001.
- [LC01] B. Leuf and W. Cunningham. *The Wiki Way: Collaboration and Sharing on the Internet*. Addison-Wesley, 2001.
- [Lev02] S. Levy. Living in the blog-osphere. *Newsweek*, 2002.
- [Lie99] J. Liebowitz, editor. *Knowledge Management Handbook*. CRC Press, 1999.
- [Lit02] M. Litchfield. ...but presses must stress ideas, not markets. *The Chronicle of Higher Education*, June 28 2002.
- [LJ80] G. Lakoff and M. Johnson. *Metaphors We Live By*. University of Chicago Press, 1980.

- [LP99] E. Lesser and L. Prusak. Communities of practice, social capital and organizational knowledge. *Information Systems Review*, 1(1), 1999.
- [LS01] E. L. Lesser and J. Storck. Communities of practice and organizational performance. *IBM Systems Journal*, 40(4), 2001.
- [LW90] J. Lave and E. Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, 1990.
- [Mac91] R. MacGregor. Inside the LOOM classifier. *SIGART Bulletin*, June 1991.
- [Mas02] M.K. Mason. Academic research, scholarly communication, and the serials crisis: Is it time to jump off the roller coaster of publishing for prestige?
<<http://www.moyak.com/researcher/resume/papers/var17mkm.htm>>, accessed Dec. 9, 2002.
- [McG02] D. McGuinness. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, chapter “Ontologies Come of Age”. MIT Press, 2002.
- [MF96] B. Manville and N. Foote. Harvest your workers’ knowledge. *Data-mation*, 7 July 1996.
- [Mil90] G.A. Miller. Wordnet: an online lexical database. *International Journal of Lexicography*, 3(4):235–312, 1990.
- [Mor02] P. Morville. Social network analysis.
<<http://semanticstudios.com/publications/semantics/000006.php>>, accessed Nov. 9, 2002.
- [MW02] T. Mortensen and J. Walker. Blogging thoughts: Personal publication as an online research tool. In Andrew Morrison, editor, *Researching ICTs in Context*, 2002.
- [Nar00] B. Nardi. It’s not what you know, it’s who you know. *First Monday*, 5(5), 2000.
- [NT95] I. Nonaka and H. Takeuchi. *The Knowledge-Creating Company*. Oxford University Press, 1995.
- [Odl95] A.M. Odlyzko. Tragic loss or good riddance? The impending demise of traditional scholarly journals. *Intern. J. Human-Computer Studies (formerly Intern. J. Man-Machine Studies)*, 42, 1995.

- [Odl97] A.M. Odlyzko. *Electronic Publishing '97: New Models and Opportunities*, chapter "The slow evolution of electronic publishing", pages 4–18. ICCC Press, 1997.
- [Odl00] A.M. Odlyzko. *Access to Publicly Financed Research: The Global Research Village III*, chapter "The Future of Scientific Communication", pages 273–278. NIWI, 2000.
- [Odl02] A.M. Odlyzko. The rapid evolution of scholarly communication. *Learned Publishing*, 15(2), 2002.
- [Odl03] A.M. Odlyzko. *Peer Review in Health Sciences*, chapter "Peer And Non-Peer Review". BMJ Books, 2003.
- [OEC72] *Interdisciplinarity: Problems of Teaching and Research in Universities*. OECD, 1972.
- [Orr96] J. E. Orr. *About Machines: An Ethnography of a Modern Job*. Cornell University Press, 1996.
- [P45] G. Pólya. *How To Solve It*. Princeton University Press, 1945.
- [Paq02] S. Paquet. Personal knowledge publishing and its uses in research. <<http://radio.weblogs.com/0110772/stories/2002/10/03/personalKnowledgePublishingAndItsUsesInResearch.html>>, accessed Dec. 2, 2002.
- [PB99] A. Perez and V. Benjamins. Overview of knowledge sharing and reuse components: Ontologies and problem-solving methods. In *Proc. IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5)*, Stockholm, Sweden, 1999.
- [Per02] M. Perkmann. Measuring knowledge value. <http://www.ki-network.org/downloads/knowledge_value_B7.pdf>, accessed Dec. 21, 2002.
- [PI69] J. Piaget and B. Inhelder. *The Psychology of the Child*. Basic Books, 1969.
- [Pol67] M. Polanyi. *The Tacit Dimension*. Anchor Books, 1967.
- [PP03] S. Paquet and P. Pearson. The internet topic exchange: Catalyzing self-organization in the blogosphere. submitted to the BlogTalk conference, February 2003.

- [RF02] D. Rusch-Feja. The open archives initiative and the OAI protocol for metadata harvesting: Rapidly forming a new tier in the scholarly communication infrastructure. *Learned Publishing*, 15(3), 2002.
- [Rog95] E.M. Rogers. *Diffusion of innovations (4th ed.)*. The Free Press, 1995.
- [Row02] F. Rowland. The peer-review process. *Learned Publishing*, 15(4), 2002.
- [Rue02] M. Ruef. Strong ties, weak ties and islands. *Industrial and Corporate Change*, 11(3), 2002.
- [SAA⁺00] A.Th. Schreiber, J.M. Akkermans, A.A. Anjewierden, R. de Hoog, N.R. Shadbolt, W. Van de Velde, and B.J. Wielinga. *Engineering and Managing Knowledge: the CommonKADS methodology*. MIT Press, Cambridge, Massachusetts, 2000.
- [SBF98] R. Studer, V.R. Benjamins, and D. Fensel. Knowledge engineering, principles and methods. *Data and Knowledge Engineering*, 25(1-2):161–197, 1998.
- [SDFS00] R. Studer, S. Decker, D. Fensel, and S. Staab. *Knowledge Engineering and Agent Technology*, chapter Situation and Perspective of Knowledge Engineering. IOS Press, 2000.
- [SL94] A.A. Stepanov and M. Lee. The Standard Template Library. Technical Report HP Technical Report HPL-94-34, 1994.
- [Slo99] N.J.A. Sloane. My favorite integer sequences. In C. Ding, T. Hellesteth, and H. Niederreiter, editors, *Sequences and their Applications (Proceedings of SETA '98)*, 1999.
- [SMD00] S.B. Shum, E. Motta, and J. Domingue. Scholonto: an ontology-based digital library server for research documents and discourse. *Int. J. on Digital Libraries*, 3(3):237–248, 2000.
- [Smi97] R. Smith. Peer review: Reform or revolution? time to open up the black box of peer review. *British Medical Journal*, 315, 1997.
- [SP94] F. Sillion and C. Puech. *Radiosity and Global Illumination*. Morgan Kaufmann, 1994.
- [SRKR97] B. Swartout, P. Ramesh, K. Knight, and T. Russ. Toward distributed use of large-scale ontologies. In *AAAI Symposium on Ontological Engineering*, 1997.

- [SS01] S. Buckingham Shum and T. Sumner. JIME – an interactive journal for interactive media. *First Monday*, 6(2), 2001.
- [SSSS01] S. Staab, R. Studer, H.-P. Schnurr, and Y. Sure. Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1):26–34, 2001.
- [Sve01] K.-E. Sveiby. What is knowledge management? <http://www.sveiby.com.au/KnowledgeManagement.html>, accessed Dec. 4, 2001.
- [Tho78] J. C. Thomas. A design-interpretation analysis of natural english. *International Journal of Man-Machine Studies*, 10, 1978.
- [TKE01] J. C. Thomas, W. A. Kellogg, and T. Erickson. The knowledge management puzzle: Human and social factors in knowledge management. *IBM Systems Journal*, 40(4), 2001.
- [TS98] J. Tennison and N. Shadbolt. APECKS: A tool to support living ontologies. In B. Gaines and M. Musen, editors, *Proc. 11th Workshop on Knowledge Acquisition, Modeling and Management, KAW '98*, 1998.
- [UG96] M. Uschold and M. Grüninger. Ontologies: Principles, methods, and applications. *Knowledge Engineering Review*, 11(2):93–155, 1996.
- [vHSW97] G. van Heijst, A. Schreiber, and B. Wielinga. Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 46:183–292, 1997.
- [Vic02] T. Vicsek. Complexity: The bigger picture. *Nature*, 418(131), 2002.
- [Vyg62] L. Vygotsky. *Thought and Language*. MIT Press, 1962.
- [Wal02a] D. Walker. Of Google, Amazon and weblogs: reputation management evolves. http://www.shorewalker.com/pages/reputation_manager-1.html, accessed Dec. 21, 2002.
- [Wal02b] J. Walker. Links and power: The political economy of linking on the web. In ACM Press, editor, *Proceedings of Hypertext 2002*, 2002.
- [Wel99] A. Wells. Exploring the development of the independent, electronic, scholarly journal. Master's thesis, University of Sheffield, England – Department of Information Studies, 1999.

- [Wen98] E. Wenger. Communities of practice: Learning as a social system. *Systems Thinker*, June 1998.
- [WG97] G. Wiederhold and M. Genesereth. The conceptual basis for mediation services. *IEEE Intelligent Systems*, 12(5):38–47, 1997.
- [Win03] D. Winer. The history of weblogs.
<<http://newhome.weblogs.com/historyOfWeblogs>>, accessed Jan. 23, 2003.
- [WS00] E. Wenger and Y. Snyder. Communities of practice, the organizational frontier. *Harvard Business Review*, Jan-Feb. 2000.